



**Titre:** Analyse multivariée de données d'opération d'une usine intégrée de  
pâte et papier

**Auteur:** Monica Ortiz

**Date:** 2005

**Type:** Mémoire ou thèse / Dissertation or Thesis

**Référence:** Ortiz, M. (2005). Analyse multivariée de données d'opération d'une usine intégrée  
de pâte et papier [Master's thesis, École Polytechnique de Montréal]. PolyPublie.  
Citation: <https://publications.polymtl.ca/7657/>

 **Document en libre accès dans PolyPublie**  
Open Access document in PolyPublie

**URL de PolyPublie:** <https://publications.polymtl.ca/7657/>  
PolyPublie URL:

**Directeurs de  
recherche:**  
Advisors:

**Programme:** Unspecified  
Program:

UNIVERSITÉ DE MONTRÉAL

ANALYSE MULTIVARIÉE DE DONNÉES D'OPÉRATION D'UNE USINE  
INTÉGRÉE DE PÂTE ET PAPIER

MONICA ORTIZ  
DÉPARTEMENT DE GÉNIE CHIMIQUE  
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION  
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES  
(GÉNIE CHIMIQUE)  
JUN 2005



Library and  
Archives Canada

Bibliothèque et  
Archives Canada

Published Heritage  
Branch

Direction du  
Patrimoine de l'édition

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file    Votre référence*

*ISBN: 978-0-494-16827-1*

*Our file    Notre référence*

*ISBN: 978-0-494-16827-1*

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**

UNIVERSITÉ DE MONTRÉAL  
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

ANALYSE MULTIVARIÉE DE DONNÉES D'OPÉRATION D'UNE USINE  
INTÉGRÉE DE PÂTE ET PAPIER

présenté par ORTIZ Monica

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. PERRIER Michel, Ph.D, président

M. PARIS Jean René, Ph. D, membre et directeur de recherche

M. ORCCOTOMA Jose-Antonio, Ph. D, membre et codirecteur de recherche

M. DUCHESNE Carl, Ph. D, membre

*À mes parents,  
ma sœur Olivia, mes frères Ricardo et Javier  
et à Pierre.*

## REMERCIEMENTS

Je tiens à remercier les nombreuses personnes qui m'ont aidée pendant la réalisation de ce projet qui a été pour moi une expérience très intéressante et enrichissante:

Mon directeur de recherche Jean Paris qui m'a accueillie au sein du Centre de Recherche en Ingénierie du Papier (CRIP) et mon codirecteur Jose-Antonio Orccotoma qui en a fait de même chez Paprican. Ils m'ont guidée dans les étapes importantes du projet.

Le personnel de Paprican, tout particulièrement Andrea Hagedorn pour ses conseils ainsi que Françoise Forel, Alain Roche et Ingunn Omholt pour avoir partagé avec moi leurs connaissances sur le vaste domaine de pâte et papier.

Le personnel de Bowater Gatineau et en particulier Jason Baril, Bernard Bégin et Roger Brouillette pour leur précieuse collaboration.

Louise Bernier, pour son efficacité et sa gentillesse.

Ma famille qui a été toujours présente avec tout le soutien dont j'avais besoin.

La famille Sauriol pour m'accueillir comme un de ses membres.

Magda, Zoé et Alain pour leur amitié et m'avoir fait connaître Montréal et la beauté de l'hiver québécois (même à -20° C).

Mes compagnons actuels et passés du pavillon Aisenstadt et de Poly: Alfa, Andrea, Christian, David, Erika, Fero, Fernando, Guillermo, Ian, Julie, Lionel, Xevi.

Tous ceux qui ont fait de mon séjour à Montréal une très belle expérience.

## RÉSUMÉ

Le principal souci dans l'industrie de papier journal est de produire un papier avec de bonnes propriétés pour en faciliter l'impression. La feuille doit posséder principalement deux caractéristiques: une résistance mécanique adéquate pour assurer le fonctionnement efficace des presses imprimantes et une surface appropriée pour l'impression. Dans ce travail on s'intéressait à la première caractéristique et plus particulièrement à l'étude des causes de variation de l'indice de tension du papier.

À cet effet, des données archivées par une usine intégrée de papier journal ont été récupérées pour être analysées. Ces données comportent des informations sur les propriétés des copeaux, l'opération des raffineurs, la qualité de la pâte, l'opération de la machine à papier et les propriétés du papier produit. Plusieurs techniques statistiques d'analyse multivariée ont été envisagées afin d'obtenir le maximum d'information possible à partir d'une base de données aussi large (plus de 80 variables qui ont été suivies pendant 8 mois d'opération). La technique la plus appropriée a été celle qui concerne l'obtention de modèles de projections de structures latentes, appelés modèles PLS multi-bloc hiérarchisés.

La méthode de mise en blocs hiérarchisés a permis d'identifier et de quantifier, d'une façon systématique, le rôle des étapes du procédé et des produits intermédiaires sur

l'indice de tension sens machine (SM) du papier. Cette méthode s'est révélée simple à mettre en œuvre et efficace pour l'identification des corrélations entre les variables.

Il a été trouvé que la composition de la pâte alimentée à la machine est le facteur qui a l'influence la plus importante sur l'indice de tension SM. Une augmentation du pourcentage de pâte thermomécanique produit un papier avec un indice de tension SM plus élevé, alors qu'une augmentation du pourcentage de pâte désencrée produit l'effet contraire. La longueur des fibres de la pâte thermomécanique a aussi une influence très significative sur l'indice de tension SM: des fibres plus longues produisent un papier plus fort en termes de tension SM. Les variables d'opération de la machine à papier qui présentent une influence significative sur l'indice de tension SM sont reliées aux effets de la composition de la pâte.

Finalement cette étude nous a permis d'appliquer, pour la première fois dans le domaine de pâte et papier, la méthode de mise en blocs hiérarchisés. Cette méthode était facile d'application et efficace pour l'identification et la quantification du rôle de chaque étape du procédé par rapport à la variable de réponse choisie. Les résultats de l'analyse ont mis également en évidence le caractère dynamique des procédés étudiés et donc le développement d'un modèle qui s'adapte aux changements dans le procédé devrait être favorisé.



## ABSTRACT

The main concern in the newsprint industry is the production of good quality paper for printing purposes. The sheet must have two important properties: an adequate mechanical resistance to ensure the effective operation of the presses (paper runnability) and a suitable surface for printing (printability). In this work, we were primarily interested in the first property and specifically focused on the study of the parameters causing variations in the tensile index.

For this purpose, data from an integrated newsprint mill was recovered to be analyzed. The database used comprised information about wood furnish properties, refining operation, pulp quality, paper machine operation and paper quality. Several multivariate analysis techniques were considered to obtain as much information as possible from such a broad data base (more than 80 variables which were monitored for 8 months of operation). The hierarchical blocking method combined with PLS models used in this work was found to be well-suited for this type of application.

It was found that the composition of the pulp feedstock is the most influent factor on the tensile index in the machine direction (MD). A higher MD tensile index is obtained by increasing the thermomechanical pulp fraction, whereas an increase in the fraction of deinked pulp produces the opposite effect. The thermomechanical pulp fibre length has also a very significant influence on the MD tensile index: longer fibres produce

paper with higher MD tensile. The paper machine operation variables having a significant impact on the MD tensile index are associated with the effects of the pulp composition.

Finally, in this work, the hierarchical blocking method combined with PLS models was applied for the first time in a pulp and paper context. This method was efficient and simple to identify and to quantify the influence of each process operation with respect to the chosen response variable. The results from the analysis have shown the dynamic behavior of the studied processes, thus a model which can adapt to changes in the process should be considered for future applications.

## TABLE DES MATIÈRES

DÉDICACE.....	iv
REMERCIEMENTS.....	v
RÉSUMÉ.....	vi
ABSTRACT.....	viii
TABLE DES MATIÈRES.....	x
LISTE DE TABLEAUX.....	xiv
LISTE DES FIGURES.....	xvii
LISTE DES ANNEXES.....	xx
LISTE DES SYMBOLES ET ABRÉVIATIONS.....	xxi
<b>CHAPITRE 1 : INTRODUCTION.....</b>	<b>1</b>
1.1 Problématique.....	1
1.2 Objectifs de l'étude.....	2
1.3 Organisation du mémoire.....	3
<b>CHAPITRE 2 : CADRE DE L'ÉTUDE.....</b>	<b>4</b>
2.1 Présentation de l'industrie papetière.....	4
2.2 La composition du papier.....	4
2.3 Présentation du procédé de mise en pâte thermomécanique (PTM).....	5
2.4 Propriétés de la pâte et leur détermination.....	10

2.5 Présentation du procédé de fabrication du papier journal.....	13
2.5.1 Dosage et mélange de constituants.....	13
2.5.2 Opération de la machine à papier.....	15
2.6 Propriétés du papier et leur détermination.....	21
<b>CHAPITRE 3 : REVUE DE LITTÉRATURE.....</b>	<b>24</b>
3.1 Les méthodes par projections PCA et PLS.....	24
3.2 Quelques applications des méthodes PCA et PLS.....	27
3.2.1 Applications visant le contrôle adaptif du procédé.....	27
3.2.2 Cartes multivariées de contrôle.....	30
3.2.3 Analyse du procédé par blocs.....	34
3.3 Les méthodes PCA et PLS dans l'industrie papetière.....	36
<b>CHAPITRE 4 : PRINCIPES DE L'ANALYSE EN COMPOSANTES</b>	
<b>    PRINCIPALES (PCA) ET LES PROJECTIONS DE</b>	
<b>    STRUCTURES LATENTES (PLS).....</b>	<b>42</b>
4.1 Les méthodes par projections.....	42
4.2 Définitions.....	43
4.3 Prétraitement des données.....	45
4.3.1 Mise en échelle et centrage.....	45
4.3.2 Traitement des observations hors limites.....	47
4.3.3 Transformations.....	48
4.3.4 Traitement des observations manquantes.....	48

4.4 L'analyse en composantes principales PCA.....	49
4.4.1 Interprétation géométrique de PCA.....	50
4.4.2 Interprétation algébrique de PCA.....	54
4.4.3 Détermination du nombre de composantes principales dans un modèle PCA.....	57
4.5 Les projections de structures latentes PLS.....	60
4.5.1 Interprétation géométrique de PLS.....	61
4.5.2 Interprétation algébrique de PLS.....	65
4.5.3 Détermination du nombre de composantes principales dans un modèle PLS.....	68
 <b>CHAPITRE 5 : ANALYSE DE LA VARIABILITÉ DE LA FORCE</b>	
<b>DU PAPIER DANS UNE USINE DE PAPIER JOURNAL.....</b>	
5.1 Présentation.....	70
5.2 Introduction.....	71
5.3 L'usine.....	72
5.4 Données descriptives.....	72
5.5 Méthode.....	73
5.5.1 Démarche de modélisation.....	73
5.5.2 Prétraitement des données.....	77
5.5.2.1 Synchronisation.....	77
5.5.2.2 Élimination des observations suspectes ou hors limites.....	77

5.5.2.3 Élimination de la dépendance du poids de base.....	77
5.5.2.4 Modèles pour un seul poids de base.....	78
5.6 Résultats.....	78
5.6.1 Modélisation du procédé de mise en pâte thermomécanique.....	81
5.6.2 Modélisation du procédé de fabrication du papier.....	86
5.6.2.1 Influence des propriétés de la pâte thermomécanique.....	89
5.6.2.2 Influence des propriétés du système d'approche.....	90
5.6.2.3 Influence de l'opération de la machine à papier.....	94
5.7 Conclusions.....	97
5.8 Remerciements.....	98
5.9 Références.....	99
5.10 Notes additionnelles.....	101
<b>CHAPITRE 6 : CONCLUSIONS.....</b>	<b>102</b>
6.1 Conclusions des résultats obtenus.....	102
6.2 Recommandations et perspectives.....	103
<b>RÉFÉRENCES.....</b>	<b>105</b>
<b>ANNEXES.....</b>	<b>118</b>

## LISTE DES TABLEAUX

Tableau 2.1 : Variables qui affectent le raffinage des copeaux.....	8
Tableau 2.2 : Propriétés physiques de la pâte disponibles pour l'analyse de données et leurs méthodes de mesure.....	11
Tableau 4.1 : Algorithme NIPALS PCA.....	56
Tableau 4.2 : Algorithme NIPALS PLS.....	69
Tableau 5.1 : Variables disponibles pour la modélisation de la tension SM du papier journal.....	74
Tableau 5.2 : Valeurs VIP des blocs et poids $w^*$ des variables les plus significatives du procédé de fabrication du papier.....	88
Tableau B.1.1.1 : Coefficients $R^2X$ , $R^2Y$ et $Q^2$ du modèle PLS multi-bloc et des sous modèles.....	119
Tableau B.1.1.2 : Coefficients Centrés – Réduits et Valeurs VIP du modèle PLS multi-bloc.....	121
Tableau B.1.1.3 : Coefficients Centrés – Réduits et Valeurs VIP des sous – modèles.....	122
Tableau B.1.2.1 : Coefficients $R^2X$ , $R^2Y$ et $Q^2$ du modèle PLS multi-bloc et des sous modèles.....	124
Tableau B.1.2.2 : Coefficients Centrés – Réduits et Valeurs VIP du modèle PLS multi-bloc.....	125

Tableau B.1.2.3 : Coefficients Centrés – Réduits et Valeurs VIP des sous - modèles.....	126
Tableau B.1.3.1 : Coefficients $R^2X$ , $R^2Y$ et $Q^2$ du modèle PLS multi-bloc et des sous modèles.....	127
Tableau B.1.3.2 : Coefficients Centrés – Réduits et Valeurs VIP du modèle PLS multi-bloc.....	128
Tableau B.1.3.3 : Coefficients Centrés – Réduits et Valeurs VIP des sous - modèles.....	128
Tableau B.1.4.1 : Coefficients $R^2X$ , $R^2Y$ et $Q^2$ du modèle PLS multi-bloc et des sous modèles.....	129
Tableau B.1.4.2 : Coefficients Centrés – Réduits et Valeurs VIP du modèle PLS multi-bloc.....	130
Tableau B.1.4.3 : Coefficients Centrés – Réduits et Valeurs VIP des sous - modèles.....	130
Tableau B.2.1.1 : Coefficients $R^2X$ , $R^2Y$ et $Q^2$ du modèle PLS multi-bloc et des sous modèles.....	131
Tableau B.2.1.2 : Coefficients Centrés – Réduits et Valeurs VIP du modèle PLS multi-bloc.....	133
Tableau B.2.1.3 : Coefficients Centrés – Réduits et Valeurs VIP des sous – modèles.....	134
Tableau B.2.2.1 : Coefficients $R^2X$ , $R^2Y$ et $Q^2$ du modèle PLS multi-bloc et des sous modèles.....	136



Tableau B.2.2.2 : Coefficients Centrés – Réduits et Valeurs VIP du modèle PLS multi-bloc.....	137
Tableau B.2.2.3 : Coefficients Centrés – Réduits et Valeurs VIP des sous - modèles.....	138
Tableau B.2.3.1 : Coefficients $R^2X$ , $R^2Y$ et $Q^2$ du modèle PLS multi-bloc et des sous modèles.....	139
Tableau B.2.3.2 : Coefficients Centrés – Réduits et Valeurs VIP du modèle PLS multi-bloc.....	139
Tableau B.2.3.3 : Coefficients Centrés – Réduits et Valeurs VIP des sous - modèles.....	140

## LISTE DE FIGURES

Figure 2.1 : Schéma généralisé du procédé typique de mise en pâte thermomécanique.....	7
Figure 2.2 : Schéma du procédé de fabrication de papier journal.....	13
Figure 2.3 : Schéma simplifié du système d'approche.....	14
Figure 2.4 : Schéma d'une section Fourdrinier avec ses éléments principaux.....	16
Figure 2.5 : Schéma d'un formeur supérieur.....	17
Figure 2.6 : Emplacement du dernier tirage ouvert dans une machine à papier moderne typique.....	19
Figure 4.1 : Représentation d'une matrice de données de K variables et N observations.....	43
Figure 4.2 : Représentation des matrices des variables explicatives (X) et de réponse (Y).....	44
Figure 4.3 : Illustration des opérations les plus courantes de mise à l'échelle.....	46
Figure 4.4 : Représentation graphique d'une composante PCA.....	50
Figure 4.5 : Représentation géométrique d'une composante pour $K = 3$ .....	51
Figure 4.6 : Visualisation graphique des deux composantes principales pour K égal à 3.....	52
Figure 4.7 : Représentation graphique de la décomposition de la matrice X en composantes.....	54

Figure 4.8 : Représentation graphique d'un modèle PCA.....	55
Figure 4.9 : Représentation géométrique d'un ensemble des données centrées – réduites dans les espaces de X et Y.....	61
Figure 4.10 : Représentation géométrique d'une composante PLS pour $K = M = 3$ .....	62
Figure 4.11 : Relation interne entre les valeurs des composantes $t_a$ et $u_a$ .....	63
Figure 4.12 : Représentation géométrique des deux composantes PLS pour $K = M = 3$ .....	63
Figure 4.13 : Représentation graphique d'un modèle PLS.....	65
Figure 5.1 : Exemple de mise en blocs et modélisation du procédé de mise en pâte...	75
Figure 5.2 : Exemple de mise en blocs et modélisation du procédé de fabrication de papier journal.....	76
Figure 5.3 : Élimination de la dépendance de la tension du papier par rapport au poids de base.....	79
Figure 5.4 : Série chronologique des mesures de l'indice de tension SM du papier....	80
Figure 5.5 : Histogramme de la base de données divisée en deux périodes.....	80
Figure 5.6 : Séries chronologiques de l'indice de tension mesuré et prédit des formettes.....	82
Figure 5.7 : Valeurs VIP des blocs des modèles du procédé de mise en pâte.....	83
Figure 5.8 : Poids $w^*$ des variables concernant la qualité de la pâte de chaque ligne de raffinage.....	84
Figure 5.9 : Longueur des fibres de la pâte de la ligne C vs indice de tension des formettes (1ère période).....	85

Figure 5.10 : Séries chronologiques de l'indice de tension mesuré et prédit du papier.....	87
Figure 5.11 : Longueur des fibres de la PTM de la ligne C vs l'indice de tension SM du papier (1ère période).....	89
Figure 5.12 : Séries chronologiques de l'indice de tension des formettes de PTM et DES (1ère période).....	90
Figure 5.13 : Influence de la composition de la pâte du circuit de tête sur la force du papier de la première période.....	91
Figure 5.14 : Séries chronologiques de l'indice de tension SM mesuré et prédit par le modèle de la 1ère période.....	94

## LISTE DES ANNEXES

ANNEXE A. Test de Significance du Coefficient de Détermination.....	118
ANNEXE B. Coefficients des modèles obtenus.....	119
B.1 Première Période (Juin à Septembre 2003).....	119
B.1.1 Modèles ayant comme variable de réponse l'indice de tension des formettes.....	119
B.1.1 Modèles ayant comme variable de réponse l'indice de tension SM du papier.....	124
B.1.3 Modèles ayant comme variable de réponse le degré de blancheur du papier.....	127
B.1.4 Modèles ayant comme variable de réponse le débit total de vapeur.....	129
B.2 Deuxième Période (Octobre 2003 à Janvier 2004).....	131
B.2.1 Modèles ayant comme variable de réponse l'indice de tension des formettes.....	131
B.2.2 Modèles ayant comme variable de réponse l'indice de tension SM du papier.....	136
B.2.3 Modèles ayant comme variable de réponse la consistance de la caisse d'arrivée.....	139

## LISTE DES SYMBOLS ET ABRÉVIATIONS

### SIGLES ET ACRONYMES

APPITA	Australian Pulp and Paper Industry Technical Association
ASTM	American Society for Testing and Materials
BPBMA	British Paper and Board Makers Association
CPPA-TS	Canadian Pulp and Paper Association – Technical Section
CR	Centré-réduit
CSF	Canadian Standard Freeness
CUSUM	Somme Cumulative (Cumulative Sum)
CV	Coefficient de variation
DES	Pâte Désencrée
ES	Énergie Spécifique
EWMA	Moyenne mobile exponentiellement pondérée (Exponentially Weighted Moving Average)
KPCA	Méthode PCA mise en œuvre avec l’algorithme <i>kernel</i>
KPLS	Méthode PLS mise en œuvre avec l’algorithme <i>kernel</i>
LA	Ligne de raffinage principale A
LB	Ligne de raffinage principale B
LC	Ligne de raffinage de rejets
PA	Raffineur primaire de la ligne A
PB	Raffineur primaire de la ligne B

PC	Raffineur primaire de la ligne C
PARAFAC	Analyse de facteurs parallèles (Parallel Factor Analysis)
PCA	Analyse en Composantes Principales (Principal Components Analysis)
PCR	Régression de Composantes Principales (Principal Components Regression)
PLS	Projections de Structures Latentes (Projections to Latent Structures)
PLS-DA	PLS Discriminante
PQM	Pulp Quality Monitor
PR	Production
PRESS	Somme des carrés des différences entre les valeurs réelles et les valeurs prédites (Predictive Residual Sum of Squares)
PRM	Pâte de Raffineur Mécanique
PTM	Pâte Thermomécanique
TAPPI	Technical Association of Pulp and Paper Industry
SA	Raffineur secondaire de la ligne A
SB	Raffineur secondaire de la ligne B
SC	Raffineur secondaire de la ligne C
SCAN	Scandinavian Pulp, Paper and Board Testing Committee.
SM	Sens Machine (vers le sens de la feuille sortant de la machine à papier)
SPE	Carré de l'erreur de prédiction (Squared Prediction Error)
ST	Sens Travers (vers le sens perpendiculaire à celui de la feuille sortant de la machine à papier)
TEA	Énergie à la rupture (Tensile Energy Absorption)

VC	Validation Croisée
VIP	Influence de la variable sur la projection (Variable Influence on Projection)

## SYMBOLS MATHÉMATIQUES

A	Nombre total de composantes principales (dimensions).
a	Indice rapportant à la $a^{\text{ième}}$ composante
B	Coefficients de régression d'un modèle PLS
C	Matrice des poids c
$\mathbf{c}_a$	Vecteur des poids de la $a^{\text{ième}}$ composante extraite de la matrice Y
$c_{ak}$	Poids de la $k^{\text{ième}}$ variable dans la $a^{\text{ième}}$ composante extraite de la matrice Y
E	Matrice des résidus de X
F	Matrice des résidus de Y
i	Indice rapportant à la $i^{\text{ième}}$ observation
K	Nombre de colonnes (variables) de la matrice X
k	Indice rapportant à la $k^{\text{ième}}$ variable
l	Indice rapportant le temps
$\lambda$	Facteur d'oubli
N	Nombre de rangées (observations) d'une matrice
M	Nombre de colonnes (variables) de la matrice Y
P	Matrice des valeurs p
$\mathbf{p}_a$	Vecteur des poids de la $a^{\text{ième}}$ composante extraite de la matrice X
$p_{ak}$	Poids de la $k^{\text{ième}}$ variable dans la $a^{\text{ième}}$ composante extraite de la matrice X



$s$	Écart type
$Q^2$	Fraction de la variation des données qui est prédite par le modèle
$r$	Rang d'une matrice
$R^2$	Coefficient de détermination
$R^2X$	Coefficient de détermination des variables de X
$R^2X$ (cum)	Coefficient de détermination cumulatif des variables de X
$R^2Y$	Coefficient de détermination du modèle PLS (entre les variables de X et Y)
$R^2Y$ (cum)	Coefficient de détermination cumulatif du modèle PLS
$s_k$	Écart type des valeurs de la variable k
$SSX$	Somme de carrées des valeurs de X (Sum of Squares)
$SSY$	Somme des carrés des valeur de Y
$T$	Matrices des valeurs t
$\mathbf{t}_a$	Vecteur des valeurs de la $a^{\text{ième}}$ composante extraite de la matrice X
$\mathbf{t}_{ai}$	Valeur de la l'observation i par rapport à la $a^{\text{ième}}$ composante extraite de la matrice X
$U$	Matrice des valeurs u
$\mathbf{u}_a$	Vecteur des valeurs de la $a^{\text{ième}}$ composante extraite de la matrice Y
$\mathbf{u}_{ai}$	Valeur de la l'observation i par rapport à la $a^{\text{ième}}$ dimension extraite de la matrice Y
$W$	Matrice des poids w
$\mathbf{w}_a$	Vecteur des poids corrigés de la $a^{\text{ième}}$ composante extraite de la matrice X dans un modèle PLS

$\mathbf{w}_a^*$	Vecteur des poids reliés aux coefficients de régression de la $a^{\text{ième}}$ composante extraite de la matrice X dans un modèle PLS
$w_{ak}$	Poids corrigé de la $k^{\text{ième}}$ variable dans la $a^{\text{ième}}$ composante extraite de la matrice X dans un modèle PLS
$w_{ak}^*$	Poids relié aux coefficients de régression de la $k^{\text{ième}}$ variable dans la $a^{\text{ième}}$ composante extraite de la matrice X dans un modèle PLS
$X$	Matrice des données ou des variables explicatives
$\hat{X}$	Matrice des valeurs de X prédites par le modèle.
$\bar{x}'$	Vecteur des valeurs moyennes des variables de X
$Y$	Matrice des variables de réponse
$\bar{y}'$	Vecteur des valeurs moyennes des variables de Y

## CHAPITRE 1

### INTRODUCTION

#### 1.1 Problématique

Actuellement, les usines canadiennes font face à des demandes de plus en plus exigeantes sur la qualité du papier. La compétition dans le marché d'exportations de pâte et de papier est féroce. L'industrie canadienne doit continuer à offrir des produits de bonne qualité à un prix compétitif pour maintenir sa place d'excellence parmi les producteurs mondiaux de pâte et papier.

En outre, quand le produit final ne répond pas aux besoins spécifiques du client, dans la plupart des cas, la production sera vendue à un prix inférieur, ce qui fait baisser la rentabilité de l'usine. Le principal souci dans l'industrie de papier journal, est le comportement du papier lors de l'impression. La feuille doit posséder principalement deux caractéristiques: une résistance mécanique adéquate pour assurer le fonctionnement efficace des presses imprimantes et une surface appropriée pour l'impression. Dans ce travail on s'intéressait à la première caractéristique: l'étude des causes de variation des propriétés mécaniques qui affectent le comportement du papier dans les presses à impression.

À cet effet, on a utilisé des données concernant les propriétés des copeaux, de la pâte et du papier produits, ainsi que les paramètres d'opération des raffineurs et d'une machine à papier. Il fallait donc trouver la méthode d'analyse multivariable qui permet d'obtenir

le maximum d'information possible de cet ensemble de données qui compte plus de 80 variables qui ont été suivies pendant 8 mois d'opération. Les méthodes par projections, qui peuvent être appliquées pour n'importe quel type d'ensemble de données, complet ou incomplet, qui peut contenir des bruits et des variables colinéaires et où le nombre de variables peut être différent du nombre d'observations, semblaient les plus appropriés pour réaliser la tâche.

## **1.2 Objectifs de l'étude**

Le but de cette étude était d'identifier les principaux paramètres causes de variations des propriétés mécaniques qui affectent le comportement du papier dans les presses à impression. La variabilité de la tension du papier dans le sens de la machine a été choisie comme objet d'étude. La réalisation de cet objectif tient compte des points suivants:

- Récupération des données concernant les propriétés de la matière première, de la pâte et du papier produits ainsi que les paramètres d'opération des procédés de mise en pâte et fabrication du papier journal.
- Prétraitement des données afin d'obtenir un ensemble approprié pour réaliser l'analyse des variations à long terme.
- Construction de modèles du procédé de mise en pâte et de fabrication de papier, faciles à interpréter pour ainsi identifier efficacement les principales sources des variations de la force du papier.

- Formulation des mesures à appliquer dans la conduite du procédé afin d'augmenter la force du papier et de réduire sa variabilité.

### **1.3 Organisation du mémoire**

Le chapitre 2 présente le cadre de l'étude. En particulier, les grandes lignes des procédés de mise en pâte et de fabrication du papier y sont présentées. Le chapitre 3 réunit une revue de littérature sur les méthodes par projections et leurs applications. Une attention spéciale est donnée aux applications dans l'industrie papetière. Le chapitre 4 présente les principes des méthodes par projections. Leur compréhension est essentielle afin de tirer tous les avantages qu'elles possèdent pour l'analyse des données et l'interprétation des résultats. Le chapitre 5 décrit l'analyse de la variabilité de la force du papier réalisée dans une usine intégrée de papier journal. La méthodologie suivie et les résultats de la démarche de modélisation sont y présentés. Ce chapitre est autonome, car il constitue une communication faite au Congrès Fibre – Papier – Carton 2005 qui sera publiée dans les comptes rendus de ce congrès. Finalement, les conclusions de l'étude et les recommandations pour la conduite du procédé, ainsi que les voies possibles à suivre pour la continuation de ce sujet de recherche sont présentées dans le chapitre 6.

## **CHAPITRE 2**

### **CADRE DE L'ÉTUDE**

#### **2.1 Présentation de l'industrie papetière**

De nos jours, il est très difficile d'imaginer une journée où on n'utiliserait pas le papier ou un de ses produits dérivés. Leur importance est indéniable. Les usages et les applications sont très diverses et la création de nouveaux produits pour usages spéciaux est en évolution constante. En outre, l'industrie des pâtes et papiers fournit de l'emploi à un grand nombre de personnes et joue un rôle primordial dans l'économie globale du Canada et des États-Unis. Seulement au Canada en 2002, l'industrie papetière a obtenu des revenus de 35,2 milliards de \$ CAN et a employé plus de 94 000 personnes (Banque de données Strategis, 2005).

#### **2.2 La composition du papier**

Le papier est majoritairement constitué de fibres cellulosiques provenant du bois, de papier recyclé et dans une moindre mesure de fibres non sylvestres. On peut considérer que le bois est constitué en moyenne, sur une base sèche, de 43% de cellulose, de 29% d'hemicellulose, de 24% de lignine et de 4% de substances extractibles. Les proportions de chaque composant varient dans les différentes espèces d'arbres.

La cellulose détermine la qualité des fibres et rend possible leur utilisation dans la fabrication du papier. Les fibres cellulosiques présentent les propriétés suivantes:

résistance élevée à la rupture par traction, flexibilité, résistance à la déformation plastique, insolubles dans l'eau, hydrophiles, large gamme de dimensions, capacité d'absorption d'adjuvants modificateurs, chimiquement stables et relativement incolores (Smook, 2001). Ces propriétés sont grandement améliorées dans le raffinage, en éliminant la membrane primaire des fibres ce qui rend possible leur hydratation. Ceci augmente leur flexibilité et leur pouvoir de liaison.

Dans les sections suivantes les grandes lignes des procédés de mise en pâte thermomécanique et de fabrication du papier sont abordées. Pour une étude plus approfondi, plusieurs références peuvent être consultés (Biermann, 1994; Deng, 1994; Smook, 2001).

### **2.3 Présentation du procédé de mise en pâte thermomécanique (PTM)**

La mise en pâte par voie thermomécanique (PTM) est une variation du procédé de mise en pâte mécanique, par lequel les copeaux de bois sont défibrés entre les plaques d'un raffineur (Pâte de Raffineur Mécanique, PRM). Dans le procédé PRM, une vis d'alimentation conduit les copeaux dans la chambre de raffinage de l'appareil. Les copeaux sont réduits progressivement en fragments de plus en plus petits, et finalement en fibres, pendant leur passage à travers la zone de raffinage. De l'eau introduite dans la chambre de raffinage sert à régler la consistance de la pâte. Pendant le raffinage, la lignine subit un ramollissement par compression et décompression, ainsi que par le frottement bois-bois et métal-bois qui s'exerce entre les plaques du raffineur. Les effets

de dévidage et de vrillage sont évidents sur les fibres obtenues. Afin d'éliminer ces effets de «latence», la pâte obtenue doit être laissée en repos pendant quelque temps dans l'eau chaude.

Depuis les années trente, on sait que le ramollissement des copeaux par la chaleur confère une meilleure résistance à la pâte. Ceci constitue l'origine du procédé PTM, qui consiste à chauffer les copeaux de bois à la vapeur avant et pendant la première étape de raffinage. La pâte ainsi obtenue contient un plus grand pourcentage de fibres longues et moins de bûchettes que la PRM. D'habitude, l'étuvage et la première étape de raffinage sont effectués sous pression et la deuxième étape est à pression atmosphérique. La figure 2.1 montre le schéma généralisé typique du procédé de mise en pâte thermomécanique.

Le tableau 2.1 fournit une liste de variables qui affectent le raffinage des copeaux. Dans le cas de la PTM, on doit aussi accorder une attention particulière à la régulation des températures au cours de l'étuvage et du raffinage. La température de raffinage doit être inférieure à 140 °C. Au dessus de 140 °C les fibres sont facilement séparées avec une faible consommation d'énergie. Cependant, les fibres ainsi détachées sont enrobées de lignine qui, lors de son refroidissement, passe à un état vitreux qui risque de nuire à une fibrillation subséquente des fibres. Entre 120 et 130 °C, la lignine se ramolli suffisamment pour permettre une bonne séparation des fibres, mais il peut se produire des ruptures dans les aspirantes extérieures des parois secondaires des fibres.



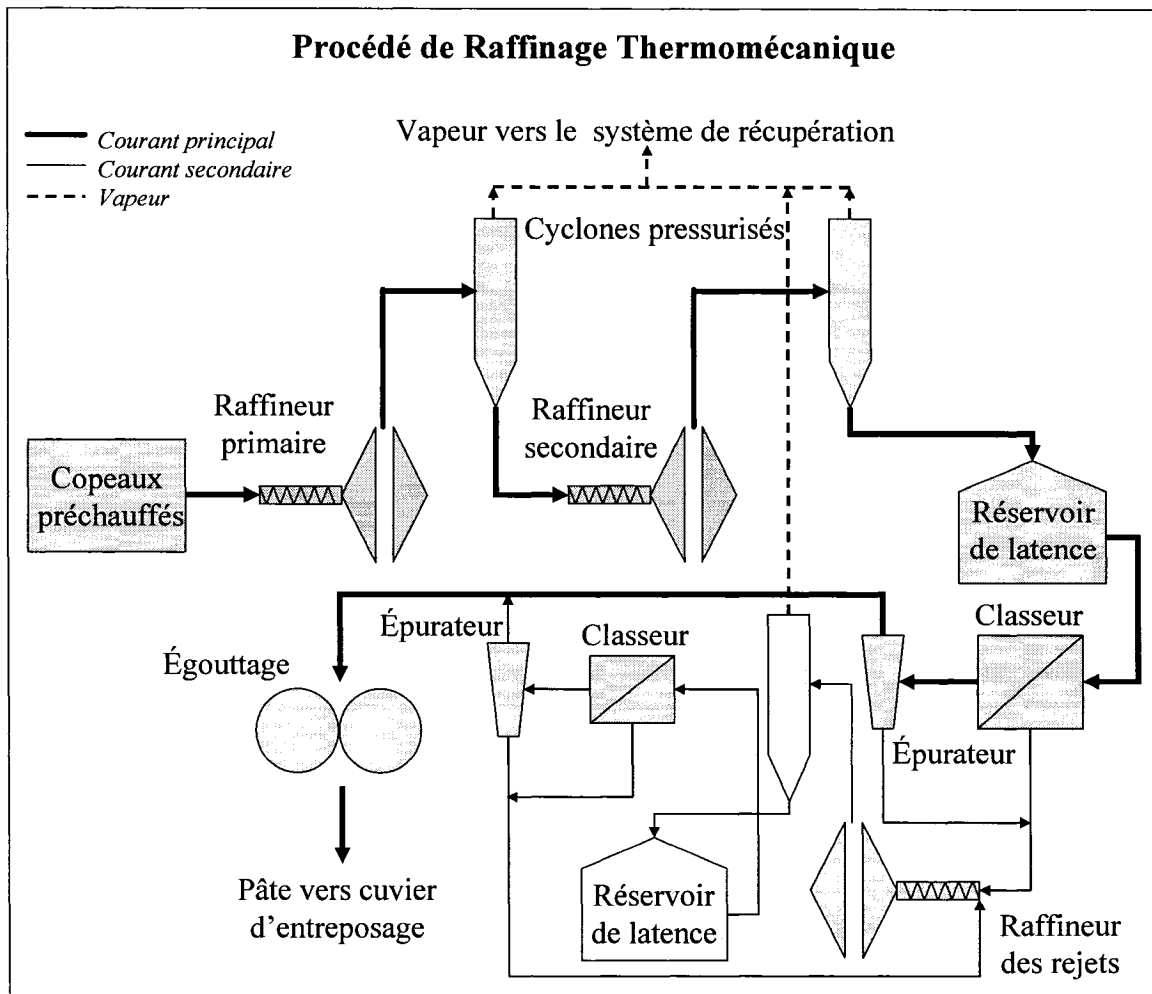


Figure 2.1 Schéma généralisé du procédé typique de mise en pâte thermomécanique.

Tableau 2.1. Variables qui affectent le raffinage des copeaux (Smook, 2001).

Matière première	Espèces, qualité générale des copeaux, degré d'humidité, absence de corps étrangers
Caractéristiques des plaques	Matériel de construction (métal), modèle, conicité, périphérie ouverte ou fermée, diamètre du disque
Conditions d'opération	Pression d'admission, température d'admission, consistance, entrefer, adjuvants chimiques, débit d'alimentation, vitesse de rotation des disques, cycle de remplacement des plaques, contre-pression à la décharge du raffineur

Quel que soit le type de pâte mécanique, la quantité d'énergie dépensée par unité de production, soit l'énergie spécifique, demeure un paramètre essentiel concernant la qualité de la pâte. Comme ce paramètre est difficile à mesurer sur une base continue, l'égouttage (la résistance présentée par les fibres à l'écoulement de l'eau) est utilisé fréquemment comme paramètre de contrôle. Plus la dépense énergétique est grande, moins l'indice d'égouttage de la pâte est élevé. Cependant, le lien entre les deux paramètres peut varier de façon importante en fonction de la qualité de la matière première (Smook, 2001). Ainsi, la pâte est fabriquée en faisant un compromis entre sa qualité et son égouttabilité sur la toile de formation de la machine à papier.

Après le raffinage, des cyclones pressurisés permettent de séparer la pâte de la vapeur. Ensuite, les fibres retrouvent leur forme allongée dans un réservoir de latence. La pâte passe ensuite dans un classeur destiné à enlever les plus gros fragments qui ont échappé à l'action des raffineurs. Les rejets de ce classeur sont généralement déchiquetés, épaissis et traités dans un autre raffineur. Le classeur sert aussi à éliminer, dans un certain degré, les éléments indésirables tels que des éclats de bois et des impuretés. Il est inefficace pour l'élimination des mini-bûchettes dont le rapport longueur/diamètre est égal à trois ainsi que les petites particules métalliques ou le sable (Chen, 1989). L'utilisation d'un système d'épuration est donc nécessaire. Les épureurs servent à purifier la pâte mais aussi à la classer. Les fibres qui sont assez développées sont envoyées dans la section d'entreposage de la pâte. Celles qui sont sous-développées sont envoyées dans la section de raffinage de rejets (Franko, 1987).

Le degré de blancheur de la pâte est affecté par plusieurs facteurs, tels que la couleur de la lignine des espèces de bois utilisées, les conditions d'entreposage des copeaux et de la pâte ainsi que l'âge de la pâte. La méthode de raffinage peut aussi modifier et intensifier la couleur associée à la lignine contenue dans la pâte. Le degré de blancheur de la pâte mécanique est fréquemment adéquat pour son utilisation finale. Cependant, dû aux facteurs mentionnés plus haut, le degré de blancheur de la pâte peut être au-dessous de celui requis pour produire du papier journal et d'autres types de papier de plus haute qualité. Dans ces cas-ci, l'opération de blanchiment est nécessaire. Afin de maintenir le rendement de la pâte, le blanchiment est fait avec des produits chimiques

qui préservent le contenu de lignine dans la pâte. Les agents de blanchiment plus utilisés sont le peroxyde d'hydrogène et l'hydrosulfite de sodium. L'hypochlorite de sodium ou potassium ainsi que le sulfite ou bisulfite de sodium sont aussi utilisés dans certains cas (Barton et al, 1987).

#### **2.4 Propriétés de la pâte et leur détermination**

Une caractérisation de la pâte produite est nécessaire afin d'optimiser le fonctionnement du procédé ainsi que les caractéristiques du produit. Elle permet aussi d'établir des relations entre les propriétés des formettes (des feuilles de papier qui sont faites au laboratoire en évacuant l'eau d'une suspension de pâte sur un moule) et les propriétés de la pâte. Divers tests ont été développés et normalisés (Technical Association of Pulp and Paper Industry [TAPPI] Test Methods) pour déterminer plusieurs propriétés de la pâte, des fibres et des formettes. Fahey (1987), Nam Law et Garceau (1989), Smook (2001), parmi d'autres sont d'excellentes références à cet égard. Ici, on se limite à mentionner, dans le tableau 2.2, les propriétés qui ont été disponibles pour faire l'analyse de la variabilité de la force du papier dans cette étude.

D'autres mesures concernant la force de la PTM, comme la tension, l'énergie à la rupture et l'élongation, ont été également disponibles. La définition de telles propriétés est présentée à la section 2.6.

Tableau 2.2 Propriétés physiques de la pâte disponibles pour l'analyse de données et leurs méthodes de mesure.

Propriété	Méthode de mesure <sup>†</sup>
Longueur des fibres	<ul style="list-style-type: none"> <li>▪ À partir d'une observation au microscope d'un nombre significatif de fibres. Toutes les fibres sont mesurées et la valeur moyenne est calculée mathématiquement.</li> <li>▪ À partir d'une grille spéciale de drainage. Une suspension fibreuse passe à travers d'une grille qui a un nombre déterminé des espaces de certaines dimensions. Le poids des fibres retenues sur la grille est mesuré et reporté comme la longueur pondérée des fibres.</li> <li>▪ Par fractionnement. On fait s'écouler à grande vitesse une suspension fibreuse diluée, parallèlement à des tamis dont les mailles sont de plus en plus petites. Au même temps, un écoulement beaucoup plus faible passe au travers des tamis. Seulement les fibres suffisamment courtes pour ne pas recouvrir les mailles peuvent passer dans la chambre suivante. L'échantillon est typiquement divisé en cinq fractions.</li> </ul>
Égouttage	<ul style="list-style-type: none"> <li>▪ La résistance opposée par les fibres à l'écoulement de l'eau est déterminée en Amérique du Nord, par le Canadian Standard Freeness Tester (CSF). L'indice CSF est le nombre de millilitres d'eau qui passent à travers l'orifice latéral de l'appareil quand la suspension s'égoutte à travers une plaque perforée.</li> </ul>

<sup>†</sup> Nam Law et Garceau (1989) et Smook (2001).

Tableau 2.2 Propriétés physiques de la pâte disponibles pour l'analyse de données et leurs méthodes de mesure (continuation).

Propriété	Méthode de mesure <sup>†</sup>
Taux de bûchettes	<p data-bbox="521 650 1406 853">Les bûchettes sont des amas de fibres de 3 à 3,5 mm de longueur, de 0,5 mm de largeur et d'environ la moitié de l'épaisseur d'une feuille de papier (0,08 mm.). Dans la feuille, elles créent des zones de faible résistance. Les principales méthodes de mesure sont:</p> <ul data-bbox="521 926 1406 1290" style="list-style-type: none"> <li data-bbox="521 926 1406 1022">▪ Mesure visuelle. Inspection de feuilles par moyen d'un microscope, projecteur ou verre illuminé.</li> <li data-bbox="521 1094 1406 1190">▪ Classeurs de laboratoire: Classeur Somerville, Classeur de mini-bûchettes PFI, Classeur Pulmac, Appareil Von Alfthan.</li> <li data-bbox="521 1262 1406 1290">▪ Détermination optique: Appareil STFI</li> </ul>

<sup>†</sup> Nam Law et Garceau (1989) et Smook (2001).

## 2.5 Présentation du procédé de fabrication du papier journal

Les grandes lignes du procédé de fabrication du papier journal, qui concerne cette étude, sont montrées dans le schéma de la figure 2.2. Dans cette section, une courte description de chaque étape est présentée, à partir de la préparation de la pâte.

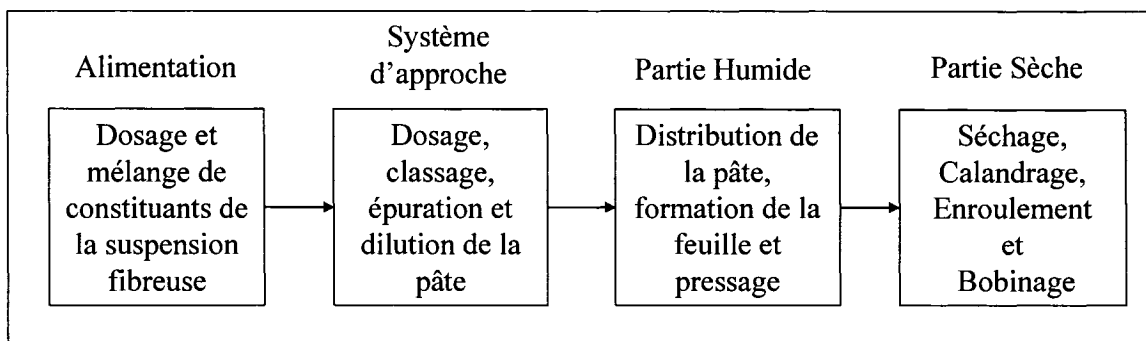


Figure 2.2 Schéma du procédé de fabrication de papier journal.

### 2.5.1 Dosage et mélange des constituants

Afin d'obtenir un mélange de pâte dont les proportions sont appropriées, il est essentiel de bien régler la concentration et le débit de chaque constituant de la suspension. Subséquemment, il est important de bien mélanger la pâte, afin de l'uniformiser, avant de l'acheminer vers la machine à papier. Ainsi, les différentes sortes de pâte (pâte thermomécanique, pâte désencrée et cassés), ainsi que des additifs destinés à améliorer les propriétés du papier, sont intégrés dans un réservoir de mélange. La pâte uniformisée dont la consistance oscille entre 2.8 et 3.2%, est ensuite conduite au cuvier de tête de la machine à papier. Ce cuvier, le premier élément du système d'approche, contient généralement le mélange final allant à la machine à papier, bien que dans certains cas, des additifs peuvent être ajoutés juste avant la caisse d'arrivée.

La figure 2.3 montre un schéma simplifié du système d'approche, qui constitue la zone comprise entre le cuvier de tête de la machine à papier et la caisse d'arrivée. Dans cette zone, on fait les opérations de dosage, classage, épuration et dilution de la suspension qui finalement sera alimentée à la caisse d'arrivée. Initialement, la suspension du cuvier de tête de la machine passe dans une caisse à niveau constant (le cuvier de pâte) qui alimente la pompe de mélange. Cette pompe sert à mélanger la pâte avec des eaux blanches de dilution (qui proviennent de la fosse sous toile de la machine à papier) et à acheminer la suspension vers la caisse d'arrivée. La consistance de la pâte sortant de la pompe est d'ordinaire entre 0,5 et 1,0 % (Smook, 1997).

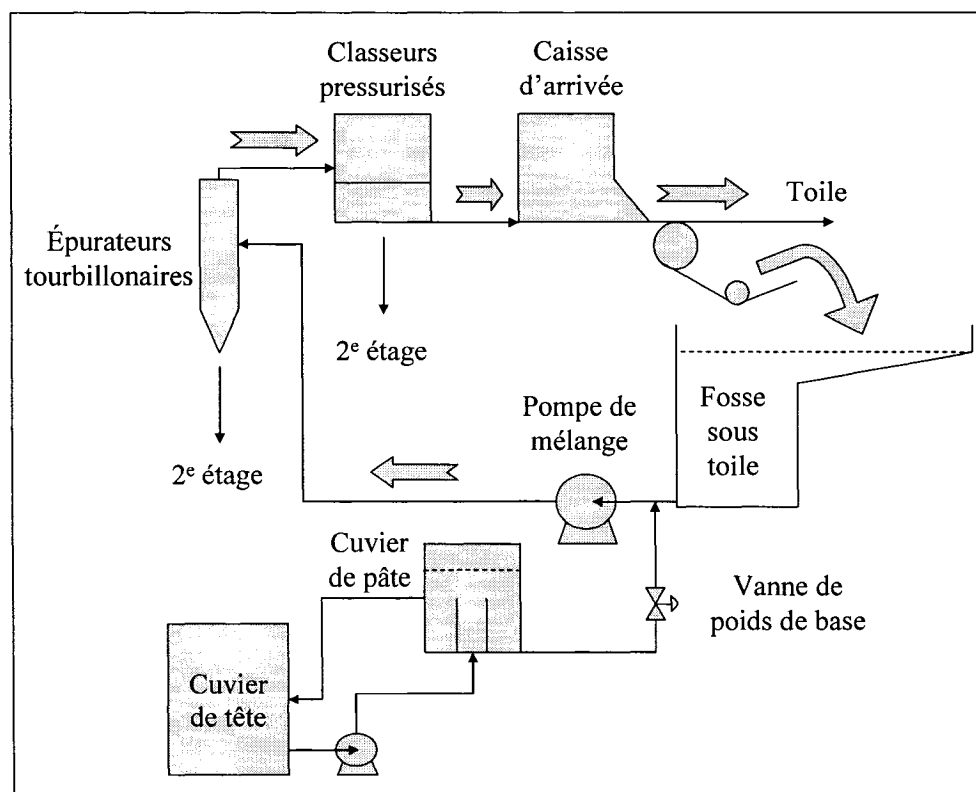


Figure 2.3 Schéma simplifié du système d'approche  
d'une machine à papier (Smook, 1997).



Bien que la suspension alimentant la machine soit épurée de façon raisonnable, des classeurs et des épurateurs sont installés en amont de la plupart des machines à papier pour éviter la présence des contaminants externes. En pratique, les classeurs éliminent les grosses impuretés et font la défloculation de la suspension, alors que les épurateurs sont conçus pour éliminer les bûchettes et les éclats de bois.

### **2.5.2 Opération de la machine à papier**

La caisse d'arrivée reçoit la suspension fournie par la pompe de mélange et la distribue sur la largeur de la machine à papier. Cette opération doit être effectuée à vitesse constante et de façon uniforme. L'ouverture des lèvres et la pression de la caisse d'arrivée sont ajustées en fonction du débit et du poids de base désiré. Ainsi, le fonctionnement de la caisse d'arrivée est un point critique de l'opération car la formation et l'uniformité de la feuille finale dépendent de la qualité de la dispersion des fibres.

Le jet de pâte qui sort de la caisse d'arrivée est dirigé vers une toile horizontale (dans le cas d'une machine de type Fourdrinier) ou entre deux toiles (dans le cas des machines à double-toile). La formation de la feuille commence ici en faisant intervenir en plus du processus d'égouttage, des effets comme la génération et l'atténuation de la turbulence, la formation du réseau fibreux, la rétention et le transport de fines particules dans le matelas fibreux, le compactage du matelas, et enfin des forces de cisaillement entre le matelas et la suspension libre (Smook, 2001).

La figure 2.4 est un schéma d'une section de Fourdrinier avec ses éléments principaux. La toile est une bande sans fin, finement tissée, qui se déplace entre deux gros rouleaux: le rouleau de tête, placé près de la caisse d'arrivée et le cylindre aspirant à l'autre extrémité. Le rouleau de tête sert à soutenir la toile sur la machine et, dans certaines configurations, il joue un rôle de formeur aspirant. Le cylindre aspirant est un cylindre creux qui contient une ou deux caisses aspirantes à vide élevé. Il sert à éliminer l'eau de la feuille. Les divers éléments entre le rouleau de tête et le cylindre aspirant sont utilisés à la fois comme support de la toile et comme organe d'égouttage (Smook, 2001).

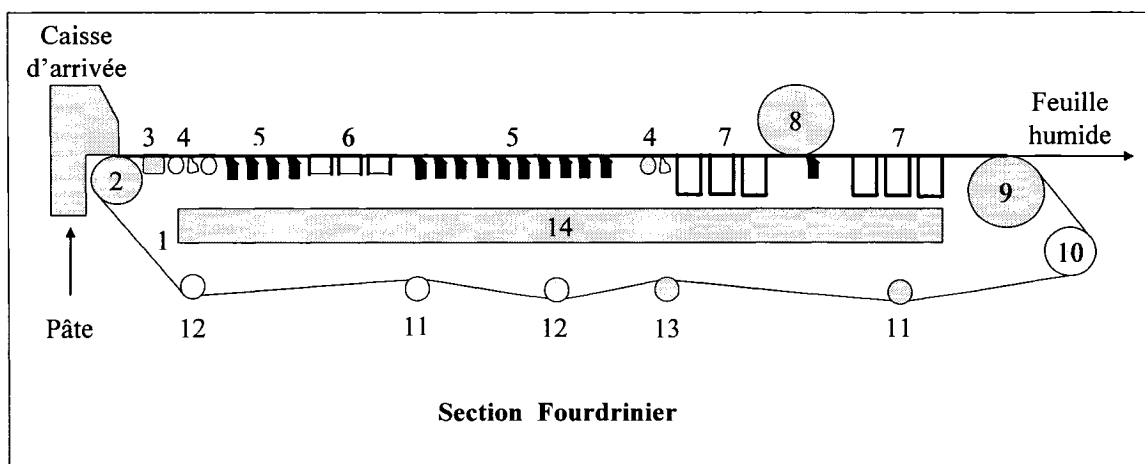


Figure 2.4 Schéma d'une section Fourdrinier avec ses éléments principaux (Smook, 2001): (1) toile métallique ou plastique, (2) rouleau de tête, (3) marbre, (4) rouleau de la table, (5) racles, (6) caisses aspirantes (partie humide), (7) caisses aspirantes (partie sèche), (8) rouleau égoutteur, (9) cylindre aspirant, (10) rouleau de retour ou moteur, (11) rouleau guide, (12) rouleau tenseur, (13) rouleau de retour, (14) fosse sous toile.

Les travaux de recherche sur d'autres méthodes de formation ont généré la prolifération de plusieurs configurations de machines à double-toile (Smook 2001). Dans les véritables formeurs à double-toile, le jet sortant de la caisse d'arrivée pénètre dans l'espace convergent entre les deux toiles. L'égouttage est produit par la pression due à la tension des deux toiles et les organes d'égouttage.

La machine à papier étudiée dans ce travail, possède un formeur Fourdrinier modifié par l'ajout d'un formeur supérieur (top-former), tel que montré à la figure 2.5. Il s'agit d'une caisse d'arrivée et un formeur Fourdrinier mis au-dessus d'un autre formeur Fourdrinier. Cette modification permet d'améliorer le processus d'égouttage et d'obtenir une feuille avec une meilleure symétrie (Hubbe, 2004).

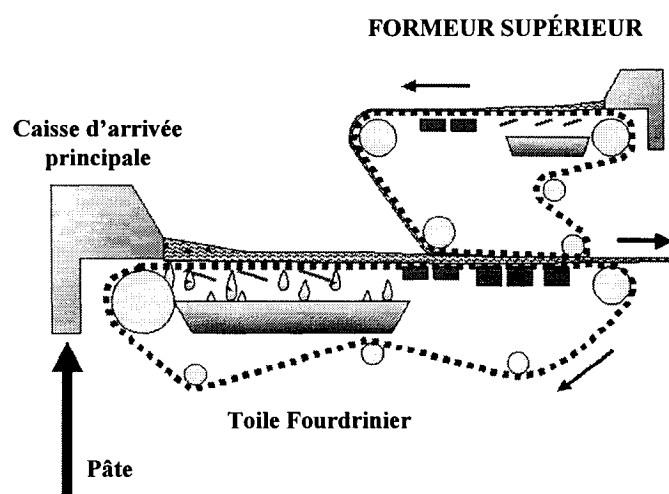


Figure 2.5 Schéma d'un formeur supérieur (Hubbe, 2004).

Le matelas fibreux sortant de la table de fabrication est transporté par une série de feutres, à travers différentes presses, afin d'extraire l'eau de la feuille et de consolider le matelas fibreux.

À l'origine, les machines étaient équipées de rouleaux leveurs à tirage ouvert pour transférer la feuille humide de la toile à la section des presses. Le transfert de la feuille était fait, sans aucun support, sur le feutre de la première presse. En maintenant une différence de vitesse entre la toile et la section des presses, on fournissait la tension nécessaire pour décoller la feuille de la toile. Cette méthode est obsolète en raison des problèmes qu'elle pose à grande vitesse. Actuellement, on décolle la feuille de la toile grâce à un feutre qui entoure un cylindre aspirant à l'endroit où se fait le contact. Avec la mise au point des rouleaux preneurs aspirants, le premier tirage ouvert se trouve dans la section des presses. Le premier tirage ouvert est toujours un point sensible car la feuille est encore assez fragile et soumise à des tensions d'origines diverses (Smook, 2001). La figure 2.6 montre l'emplacement du dernier tirage ouvert dans une machine à papier moderne typique (Pikulik, 1997).

Le pressage s'effectue dans une série de pinces, la pression augmentant d'une pince à l'autre. Il faut extraire l'eau uniformément sur toute la largeur de la machine pour que le profil d'humidité de la feuille sortant des presses et entrant en sécherie soit régulier. En outre, l'étape de consolidation de la feuille est très importante. Le contact intime

entre les fibres doit être assuré pour que les liaisons inter-fibres puissent se développer pendant le séchage (Smook, 2001).

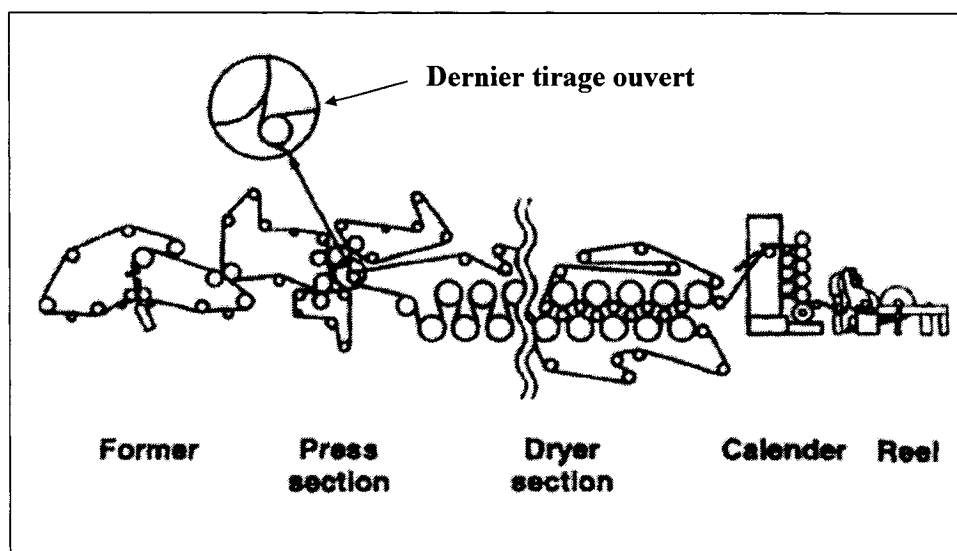


Figure 2.6 Emplacement du dernier tirage ouvert dans une machine à papier moderne typique.

L'eau qui s'écoule de la pâte pendant le processus d'égouttage est appelée l'eau blanche. Elle contient les particules non retenues dans la feuille: des fibres et des adjuvants. La majeure partie des eaux blanches provient de la toile Fourdrinier. Elle est collectée dans la fosse sous toile, pour ensuite être réutilisée directement ou à travers un système de récupération des fibres. Les eaux blanches sont réutilisées comme eau de dilution de la pâte, comme eau des rinceurs de la toile, comme eau de la section des presses, etc. La récupération des fibres des eaux blanches est faite plus couramment par

des appareils tels que les ramasse-pâte à flottation et les filtres à disques. Dans les ramasse-pâte à flottation, on récupère les fibres ainsi que d'autres solides, en fixant sur ces particules de petites bulles d'air qui les font flotter à la surface où on les récupère par écumage. Les filtres à disques sont constitués par un ensemble de disques, monté sur un axe central. L'eau s'écoule au milieu filtrant de chaque secteur, puis au centre de l'arbre rotatif. L'eau filtrée et le gâteau de filtration sont récupérés pour être réutilisés.

Le papier sortant des presses passe ensuite sur une série de cylindres sécheurs chauffés par de la vapeur vive. L'eau qui reste dans la feuille est ainsi évaporée au contact de ces cylindres et retirée par ventilation. La feuille est plaquée contre le cylindre par une toile synthétique perméable appelée feutre sécheur. Sur la plupart des machines à papier, on regroupe les cylindres dans trois ou cinq sections de feutres indépendantes dont on peut régler séparément la vitesse pour maintenir la feuille en tension entre deux sections et compenser son retrait (Smook, 2001).

La plupart des grades de papier subissent un calandrage afin d'obtenir une surface lisse pour l'impression. L'opération de calandrage la plus courante consiste à faire passer le papier dans une ou plusieurs pinces formées par un ensemble de rouleaux en fonte. La réduction d'épaisseur de la feuille et le degré de lissé sont fonction de la pression et du temps de résidence dans la pince, ainsi que du nombre de pinces.

Finalement, après l'avoir séché et calandré, il faut récupérer le papier sous une forme adaptée aux transformations ultérieures. À cet effet, la plupart de machines modernes possèdent des enrouleuses, où un tambour enrouleur est entraîné par un moteur doté d'une charge suffisante pour tendre correctement la feuille qui sort de la calandre. Une bobineuse sert à découper la feuille des bobines de papier dont la largeur correspond à celle de la machine, en bandes qui seront ré-enroulées en plus petits bobineaux. Pendant cette opération, on découpe de 3 à 5 cm sur les bords de la feuille; ces rognures sont acheminées par un convoyeur vers le tritrateur à c

assés. Les bobineaux sont emballés et expédiés au client, par exemple dans le cas du papier journal, ou bien subissent une transformation supplémentaire: aspirantage, calandrage ou découpe en feuilles.

## **2.6 Propriétés du papier et leur détermination**

La facilité de passage sur les presses dans la salle d'impression, l'imprimabilité, ainsi que l'apparence sont des propriétés importantes pour la production du papier journal. Ces propriétés sont affectées directement par la qualité du papier. Un papier plus fort aura un meilleur comportement lors de l'impression. De bonnes propriétés de surface conduisent à une bonne imprimabilité et de bonnes propriétés d'apparence donnent un meilleur aspect au papier.

Dans cette étude on s'intéresse à la facilité de passage sur les presses dans la salle d'impression et donc à la force du papier. Les définitions des propriétés du papier

reliées à la force et leur méthode de mesure sont ensuite présentées. La majorité des tests des propriétés ont été normalisés par plusieurs organismes industriels tels que Technical Association of the Pulp and Paper Industry (TAPPI), Canadian Pulp and Paper Association – Technical Section (CPPA – TS), Scandinavian Pulp, Paper and Board Testing Committee (SCAN), American Society for Testing and Materials (ASTM), British Paper and Board Makers Association (BPBMA) et Australian Pulp and Paper Industry Technical Association (APPITA).

*La résistance au déchirement* est la force moyenne requise pour continuer la déchirure du papier à partir d'une coupure initiale d'une simple feuille de papier ou de plusieurs feuilles. Cette force peut être déterminée avec l'appareil Elmendorf (Smook, 2001), qui utilise un pendule tombant qui produit la continuation de la déchirure de la feuille en appliquant une force de façon perpendiculaire au plan de la feuille. La perte d'énergie potentielle du pendule (déterminée à partir de la hauteur de ces oscillations) est reliée à la force nécessaire pour continuer la déchirure. Si la coupure a été faite dans la direction de la machine à papier, la force mesurée correspond à la résistance au déchirements sens machine (SM). Si la coupure a été fait dans l'autre sens, la force correspond à la résistance au déchirement dans le sens travers (ST).

*La tension* du papier, aussi appelée *la résistance à la rupture par traction*, est la force maximale par unité de largeur qui peut être appliquée sur une bande de papier, avant qu'elle ne se casse. Cette propriété peut être mesuré dans les deux directions (SM et



ST). On la détermine en mesurant la force nécessaire pour rompre une bande de papier de longueur normalisée. La vitesse de traction (la cadence de chargement) est également spécifiée dans la norme. *L'allongement* du papier avant rupture est déterminé simultanément. Certains appareils de mesure fournissent un diagramme de la force appliquée vs l'allongement subi par la bande de papier. L'aire sous la courbe représente *l'énergie à la rupture*, définie comme le travail nécessaire pour déchirer la bande du papier testée.

Ces propriétés sont mesurées afin de caractériser le papier mais elles sont utilisées aussi pour la caractérisation de la pâte à partir de formettes.

## **CHAPITRE 3**

### **REVUE DE LITTÉRATURE**

#### **3.1 Les méthodes par projections PCA et PLS**

Selon Jolliffe (1986), l'analyse en composantes principales PCA est probablement la technique d'analyse multivariée la plus ancienne et la mieux connue. La méthode a été introduite par Karl Pearson en 1901 mais Harold Hotelling en 1933 a été le premier à la présenter telle qu'on la connaît actuellement. L'accès généralisé aux ordinateurs ainsi que le développement de ressources informatiques, de plus en plus puissantes, a permis d'appliquer ces techniques d'analyse multivariée pour résoudre des problèmes d'une grandissante complexité. La définition et le calcul des composantes principales sont explicites et simples. Ceci est probablement la raison pour laquelle on trouve de nombreuses dérivations de la méthode en réponse à divers types de systèmes et de procédés.

Plusieurs documents concernant la théorie de l'analyse en composantes principales ont été publiés. Jackson (1980) présente une introduction au concept de composantes principales ainsi que la démarche numérique, pour faire des estimations, des tests de signifiante et l'analyse de résidus. Jolliffe (1986) et Jackson (1991) ont présenté des ouvrages dédiés entièrement à l'analyse en composantes principales, la théorie et la représentation graphique de la méthode ainsi que l'interprétation des résultats avec des nombreux exemples. Wold et al (1987) est aussi une excellente référence sur ce sujet.

Geladi et Kowalski (1986) mentionnent que le travail pionnier sur la méthode PLS a été fait en grande partie par Herman Wold à la fin des années soixante dans le domaine de l'économétrie. La base de la méthode est établie entre le début des années soixante-dix et la décennie suivante. Pendant cette période la méthode gagne en intérêt parmi les économétriciens et les sociologues et elle est appliquée premièrement dans le domaine de la chimie par les groupes de Svante Wold et Harald Martens et dans le domaine de la génie chimique par le groupe de John F. MacGregor. L'application de la méthode ne cesse de se répandre aux différents domaines en intégrant les modifications pertinentes afin de s'adapter aux besoins spécifiques de chaque discipline. Les détails historiques sur le développement de la méthode PLS peuvent être consultés dans Geladi (1988).

Il y a plusieurs algorithmes pour calculer les composantes principales, l'algorithme NIPALS (Nonlinear Iterative Partial Least Squares) et l'algorithme dit *kernel* étant les plus utilisés.

Geladi et Kowalski (1986) ont présenté un tutorial complet sur la méthode PLS et les algorithmes NIPALS PCA et NIPALS PLS. Miyashita et al (1990) proposent une modification à l'algorithme NIPALS original pour assurer la convergence du calcul de la première composante. Tenenhaus et al (1995) et Tenenhaus (1999) sont des références en français intéressantes qui exposent deux applications industrielles sur la méthode PLS ainsi que les grandes lignes de l'algorithme NIPALS avec leurs

principales propriétés. La mention de l'ouvrage de Eriksson et al (2001) qui expose clairement les méthodes PCA et PLS dans le cadre de l'analyse multivariée de données et qui fait toujours référence à l'algorithme NIPALS, est incontournable.

Concernant l'algorithme *kernel*, plusieurs travaux ont été réalisés. Par rapport à l'analyse de composantes principales, Schölkopf et al (1997) proposent une nouvelle méthode non-linéaire pour la mettre en œuvre. Ceci devient le principe fondamental de l'algorithme proposé par Lu et al (2003), considéré comme plus robuste que l'original. L'ouvrage de Schölkopf et Smola (2001) inclut un support pédagogique sur l'analyse en composantes principales avec l'algorithme *kernel*, souvent appelé méthode KPCA. En ce qui concerne la méthode PLS, l'algorithme *kernel* (KPLS) a été originellement développé par Lindgren et al (1993). Il est particulièrement efficace pour le cas des matrices où le nombre d'observations est plus grand que le nombre de variables. De Jong et Ter Braak (1994) proposent une reformulation du même algorithme afin d'éclaircir l'approche *kernel* et la méthode PLS en soit. La version de l'algorithme KPLS pour le cas des matrices où le nombre de variables est plus grand que le nombre d'observations est exposée par Rännar et al (1994, 1995). Dayal et MacGregor (1997a) présentent deux algorithmes KPLS et l'algorithme NIPALS améliorés (le code MATLAB<sup>®</sup> y compris) qui réduisent le temps de calcul ainsi que la taille de ressources informatiques requises en termes de mémoire disponible. Ceci a permis le développement d'une approche concernant le contrôle adaptatif du procédé qui sera

abordée dans la section 3.2.1. Finalement Bennett et Embrechts (2003) présentent des dérivations de l'algorithme KPLS qui sont orientées vers l'apprentissage-machine<sup>1</sup>.

La prochaine section présente quelques applications des méthodes PCA et PLS intéressantes à explorer, étant donné le potentiel d'application pour l'étude du procédé de mise en pâte et de fabrication du papier.

## **3.2 Quelques applications des méthodes PCA et PLS**

### **3.2.1 Applications visant le contrôle adaptif du procédé**

La plupart des procédés industriels sont variables dans le temps. Les changements des conditions d'opération génèrent le besoin de mettre à jour le modèle pour maintenir sa capacité descriptive et prédictive. Dayal et MacGregor (1997b) ont proposé un algorithme *kernel* pour la mise à jour exponentiellement pondérée d'un modèle PLS. Le potentiel de cette approche est illustré par deux exemples concernant la simulation du contrôle adaptif d'un réacteur agité de façon continue et la mise à jour d'un modèle prédictif d'un circuit de flottation des minéraux. En comparant les modèles obtenus avec l'algorithme KPLS récursif et l'estimation récursive par des moindres carrés, il a été démontré que les premiers donnent les meilleures prédictions.

---

<sup>1</sup> L'apprentissage-machine est un effort de recherche qui a comme but le développement d'un système expert avec la création de programmes d'ordinateur qui peuvent apprendre par l'expérience.

Qin (1993) a présenté sa version récursive de l'algorithme NIPALS modifié dont l'application illustrée est la modélisation d'un système de reformage catalytique d'une raffinerie de pétrole. Cet algorithme est la base de plusieurs approches (Qin, 1998) de modélisation du procédé en ligne et hors ligne qui sont en suite brièvement décrites.

La première approche avec l'algorithme PLS récursif (RPLS) proposé par Qin (1998) réalise la mise à jour du modèle en intégrant les nouvelles données du procédé aux valeurs des composantes de l'ancien modèle. Un nouvel ensemble de données est ainsi constitué, l'algorithme NIPALS modifié est alors mis en œuvre en fonction de ces données et un nouveau modèle actualisé est obtenu. Une deuxième approche consiste à mettre en œuvre l'algorithme NIPALS modifié afin d'obtenir un modèle PLS qui tient seulement compte des nouvelles données. Les valeurs des composantes de ce modèle sont alors intégrées à celles des composantes de l'ancien modèle pour former ce qui sera un nouvel ensemble de données. L'algorithme NIPALS modifié est à nouveau mis en œuvre avec cet ensemble et donc un nouvel modèle actualisé est obtenu. Une troisième approche, dite à fenêtre mobile, est proposée. Pour éliminer les données les plus anciennes de l'ensemble afin d'adapter le modèle aux nouvelles conditions d'opération, chaque fois qu'un nouveau ensemble de données (vu comme un bloc) est disponible et donc un nouveau bloc de composantes est obtenu, les composantes les plus anciennes sont éliminées de l'ensemble original tandis que les nouvelles y sont intégrées et un nouveau modèle PLS est obtenu. Une quatrième approche implique un facteur d'oubli  $\lambda$  ( $0 < \lambda \leq 1$ ) qui est appliqué aux anciennes composantes chaque fois

qu'un nouveau bloc de composantes est intégré à l'ensemble. De cette façon le poids de chaque bloc de composantes diminue exponentiellement. Avec une petite valeur de  $\lambda$ , l'ancienne information est oubliée plus rapidement.

Un désavantage des algorithmes de Qin est que le nombre de composantes calculées est plus grand que celui qui serait déterminé par validation croisée, ce qui fait diminuer l'efficacité de l'approche. En outre, en fonction du prétraitement des données, certains paramètres, comme les valeurs moyennes et les écarts type, doivent être mises à jour à chaque fois qu'un nouveau bloc de données est disponible afin de le mettre à l'échelle.

Le facteur d'oubli  $\lambda$  peut être choisi arbitrairement ou estimé de plusieurs manières. Vijaysai et al (2003) proposent une méthode pour calculer  $\lambda$  qui tient compte du changement de la corrélation entre les variables explicatives et les variables de réponse des blocs d'entrée et de sortie. Si les paramètres du procédé ne changent pas, la corrélation entre les variables explicatives et de réponse est presque constante et  $\lambda$  est égal à 1. Cheng et al (1988, 1993) et Cooper et Worden (2000) sont d'autres références intéressantes concernant le calcul du facteur d'oubli. Des versions récursives des algorithmes PCA et PLS sont présentés par Li et al (2000), Wang et al (2003), Capron et al (2005) et Lee et al (2005).

### 3.2.2 Cartes multivariables de contrôle

Les cartes de contrôle sont un outil qui permet de faire le suivi du procédé dans le temps, avec le but de vérifier que le procédé est sous contrôle statistique. On dit qu'un procédé est sous contrôle statistique quand les valeurs des variables suivies restent près de leur valeur cible et que les causes de la variabilité observée sont uniquement aléatoires et non assignables. Les cartes de contrôle comme la carte de Shewhart, la carte des sommes cumulées CUSUM (Cumulative Sum) et la carte de la moyenne mobile exponentiellement pondérée EWMA (Exponentially Weighted Moving Average) sont utilisées pour détecter l'apparition d'un événement dont les causes sont assignables. L'élimination de ces causes conduit à l'amélioration de la qualité du produit et de la conduite du procédé.

La définition ainsi que la construction de chacune des cartes peut être consultée dans plusieurs documents (Lamouille et al, 1989; Wheeler, 1995; Levinson, 1997; Wise et Fair, 1998). Une référence intéressante est Hunter (1986). Il présente les différences parmi les cartes nommées ci-haut, en exposant particulièrement la construction de la carte EWMA qui conduit à une équation empirique de contrôle dynamique.

Telle que conçue, une carte de contrôle fait le suivi d'une seule variable du procédé. Dans la plupart des procédés industriels modernes, beaucoup de variables (autour de 100 et plus) sont mesurées dont plusieurs sont fortement corrélées. En examinant une variable à la fois, l'interprétation et le diagnostic d'un événement anormal dans le



procédé devient très compliqué. De plus, la justesse de détection des anomalies est plus erronée au fur et à mesure que la corrélation augmente. Il est souhaitable d'extraire l'information de toutes les données disponibles pour construire un schéma de suivi et de diagnostic du comportement du procédé qui soit fonctionnel. Ceci est le point de départ du contrôle statistique multivariable du procédé où les méthodes de projections PCA et PLS peuvent être utilisées pour construire la version multivariable de chacune des cartes de contrôle. À cet égard, Jackson (1985) est une bonne référence. Il présente une analyse des techniques de contrôle de qualité multivariable, concentrée sur les cartes de contrôle ainsi que l'utilisation des composantes principales.

Une carte Shewhart univariable typique est construite à partir de la série chronologique de la variable observée, la valeur moyenne des données (idéalement la valeur cible des spécifications du procédé) et les limites de contrôle qui sont fixés à la valeur moyenne  $\pm 3$  écarts types. L'apparition d'un point au-delà de cet intervalle signale que le procédé est hors de contrôle. Ce signal ne tient pas compte de l'histoire du procédé car il dépend seulement du dernier point sur la carte de contrôle. Ainsi, la carte Shewhart n'a pas de mémoire.

La version multivariable de la carte Shewhart peut être construite à partir des valeurs des variables latentes ou des valeurs Hotelling  $T^2$ . La carte des valeurs des composantes présente l'évolution du procédé en termes d'une composante, la ligne correspondant à la valeur cible et les limites de contrôle. Cependant dans la plupart de cas, on a besoin

de plus d'une composante pour avoir une description complète du système. Il faudrait construire une carte de contrôle pour chaque composante du modèle. Une option plus attirante est présentée par la carte de valeurs Hotelling  $T^2$ . Les valeurs Hotelling  $T^2$  sont des sommes pondérées des valeurs de toutes les composantes décrivant le système. Elles donnent un indice sur la localisation des observations par rapport au centre de l'hyperplan conformé par les composantes du modèle (Eriksson et al, 2001). Ainsi, la carte construite avec les valeurs Hotelling  $T^2$  et les limites de contrôle correspondantes, permet de détecter les points hors limites qui représentent les déviations les plus importantes par rapport à la variation systématique du procédé.

Kresta et al (1991) ainsi que Nomikos et MacGregor (1995) ont proposé, dans le même ordre d'idées, l'utilisation des graphiques des valeurs latentes et du carré de l'erreur de prédiction SPE (Squared Prediction Error). Les premiers illustrent le processus de détection des fluctuations dans le procédé en utilisant des données des simulations d'un réacteur à lit fluidisé et d'une colonne de distillation extractive tandis que les deuxièmes utilisent des données d'un réacteur de polymérisation en discontinu. Dans les deux travaux, il est montré que ces cartes de contrôle multivariable peuvent être plus efficaces que leur contrepartie univariable pour la détection de changements dans le procédé ainsi que des causes de tels changements.

La carte CUSUM univariable typique est un graphique de la somme cumulative des déviations des observations par rapport à la valeur cible. Tant que la moyenne des

valeurs observées est proche de la valeur cible, les valeurs sur la carte CUSUM seront aléatoirement distribuées autour de zéro. Si la moyenne des observations change et s'écarte de la valeur cible, la pente de la ligne sur la carte de contrôle change en indiquant un déplacement de la moyenne par rapport à la valeur cible. Ainsi, la carte CUSUM possède une «mémoire d'éléphant» (Hunter, 1986) car elle tient compte de toutes les valeurs obtenues et leur donne le même poids. La carte CUSUM multivariable peut être construite à partir de la somme cumulative des valeurs des composantes et, telle que la version univariable, elle est utile pour détecter des perturbations mineures dans le développement du procédé (Eriksson, 2001).

Dans le calcul de la valeur moyenne exponentiellement pondérée EWMA, les observations concernant une variable sont pondérées en donnant plus de poids aux observations les plus récentes. Ainsi, pour deux cas extrêmes, une carte de contrôle EWMA peut être équivalente à une carte Shewhart, en donnant tout le poids à la dernière observation, ou à une carte CUSUM, en donnant le même poids à toutes les observations. Les limites de contrôle de la carte EWMA sont fixés à la valeur cible  $\pm 3$  écarts types qui sont calculés tenant compte de la constante qui détermine la mémoire de la carte.

La carte EWMA multivariable peut être produite à partir des valeurs de composantes. Ce type de carte a fait l'objet de plusieurs travaux. Lucas et Saccucci (1990) ont présenté les propriétés de la valeur moyenne exponentiellement pondérée dans le cadre

du contrôle de procédé et Lowry et al (1992) ont exposé des lignes directrices pour leur implantation. Scranton et al (1996) ont présenté des avantages de l'utilisation des composantes principales au lieu de surveiller séparément toutes les variables. Dans le travail de Martin et al (2002), l'application d'une carte EWMA construite avec les composantes principales concernant le procédé de fabrication d'un film de polymère, est présentée. Il a été montré que l'identification des perturbations et de leurs sources s'est faite d'une façon explicite.

Finalement, Lowry et Montgomery (1995), Zullo (1996), MacGregor et Kourti (1995), Wachs et Lewin (1999), Zhang et al (1999), Tsung (1999), Chen et Liao (2001), Kano et al (2001) et Kourti (2002), peuvent être consultés pour une étude plus approfondie sur les applications de ces méthodes de contrôle multivariable.

### **3.2.3 Analyse du procédé par blocs**

En dépit d'une relative simplicité, l'interprétation de résultats des modèles PLS devient compliquée lorsqu'on doit considérer un nombre important de coefficients de régression. La réduction du nombre de variables peut être une option attrayante mais on risque de perdre une partie de l'information. L'approche de mise en blocs hiérarchisés présentée par Wold et al (1996) constitue une alternative intéressante. Elle consiste à regrouper les variables dans des blocs caractéristiques, à développer un modèle PLS pour chaque bloc et à utiliser les composantes ainsi obtenues comme de nouvelles variables dans un modèle PLS global appelé modèle multi-bloc. Ainsi on peut faire la

distinction entre deux niveaux. Dans le premier niveau on trouve les sous-modèles qui englobent les relations entre les variables du procédé et dans le deuxième niveau on trouve le modèle multi-bloc qui contient les relations entre les blocs. Il est sous-entendu que tous les modèles du premier niveau ainsi que le modèle multi-bloc partagent les mêmes variables de réponse.

Kourti et al (1995) et Rännar et al (1998) ont développé l'application de mise en blocs hiérarchisés pour la surveillance adaptative de lots du produit. L'objectif est de suivre la trajectoire du procédé pour détecter et éliminer les défauts qui peuvent affecter la qualité du produit final et ensuite redéfinir les limites de contrôle. Ceci conduit à une production plus consistante.

Westerhuis et al (1998) ont comparé plusieurs versions des méthodes PCA et PLS hiérarchisés et multi-bloc. Les auteurs rapportent des problèmes de convergence des algorithmes et dans les pires cas des pertes d'information. Ils proposent certaines modifications afin de résoudre ces problèmes.

Étant donné la nature et la complexité des procédés de mise en pâte et de fabrication du papier, ainsi que la structure de l'ensemble de données disponible pour l'analyse, l'approche d'analyse du procédé par blocs possède un potentiel intéressant pour résoudre le problème posé dans cette étude. Elle a donc été choisie comme méthode de modélisation de la force du papier dans ce travail.

### **3.3 Les méthodes PCA et PLS dans l'industrie papetière**

Les méthodes par projections sont de plus en plus utilisées dans l'industrie pour la modélisation, surveillance et contrôle des procédés. Dans le cas de l'industrie de pâtes et papiers, plusieurs applications de ces méthodes ont été explorées dans les diverses parties du procédé. Wold (1994) a proposé des modifications aux algorithmes PCA et PLS pour obtenir des modèles, à partir d'observations exponentiellement pondérées. Cette démarche est illustrée avec des données concernant l'opération d'une machine à papier, ce qui a permis de prouver la capacité prédictive à court terme des modèles obtenus ainsi que leur utilité pour trouver les corrélations entre les diverses variables explicatives et de réponse.

Dayal et al (1994) ont étudié l'utilisation de réseaux neuraux et de la méthode PLS pour l'obtention de modèles empiriques concernant l'opération continue d'un lessiveur Kamyr, a fin de prédire la qualité de la pâte Kraft en termes de l'indice Kappa, ainsi que pour obtenir une meilleure compréhension des relations existantes entre les variables. Les résultats des deux méthodes sont semblables en produisant des modèles qui capturent jusqu'à 62 % de la variabilité de l'indice Kappa. Cependant, seulement les modèles PLS ont permis de mieux comprendre le procédé et d'identifier les variables de procédé qui affectent le plus l'indice Kappa et donc produisent un modèle plus fonctionnel.

Teppola et al (1998) ont obtenu des modèles PCA à partir des données concernant la partie humide d'une machine à papier. Ils ont mis en évidence l'utilité des cartes de valeurs des composantes et des poids pour la visualisation de l'état du procédé, la détection des problèmes dans le développement du procédé et les variables qui en sont responsables.

Kesavan et al (2000) ont présenté une méthode de contrôle afin de réduire la variabilité de la qualité de la pâte à la sortie d'un lessiveur discontinu. Un modèle PLS est utilisé à cet effet dont la capacité prédictive est testée expérimentalement. Le modèle PLS donne des meilleures prédictions que le modèle traditionnel du facteur H de Chari.

Ivanov (2001, 2002, 2003) a exposé l'utilisation habituelle des méthodes PCA, PLS et PLS discriminante (PLS-DA) à Tembec Inc., pour l'étude des données archivées, la surveillance du procédé et la prédiction de certaines propriétés des produits, ainsi que pour faciliter la détection de causes des changements dans l'opération et accroître les connaissances sur le procédé qui permettent de l'optimiser. Les modèles hors ligne sont utilisés pour améliorer la connaissance du procédé et identifier les causes des événements spéciaux tandis que les modèles en ligne permettent le contrôle de la production de différents grades de papier en temps réel. Deux autres applications intéressantes méritent d'être mentionnées. L'une consiste à utiliser des modèles PLS-DA pour prédire le comportement du produit lors de son utilisation finale, en comparant les nouveaux lots avec les lots antérieurs. La seconde porte sur la combinaison des

méthodes PLS-DA et PCA pour modéliser plusieurs grades de papier carton simultanément. Lane et al (2000, 2001) ont traité aussi la réduction dans le nombre de modèles pour faire le suivi de la production de plusieurs grades d'un produit.

Lane et al (2002) ont repris l'approche de Wold (1994) et ont proposé l'utilisation des modèles récurrents EWPCA (exponentially weighted PCA) dont le comportement adaptatif permet de détecter les changements dans le procédé qui sont considérés comme anormaux et donc de prendre les actions correctives nécessaires. L'application de l'approche proposée au procédé de formation de la feuille illustre le potentiel de la méthode pour faire le suivi d'un procédé dynamique. L'avantage le plus important de cette méthode est la réduction des fausses alarmes signalées par les cartes de contrôle par rapport à celles produites quand la méthode PCA conventionnelle est appliquée pour le même type de procédé.

Tessier et al (1999) ont développé un modèle PLS pour optimiser la distribution des dimensions de copeaux (longueur et épaisseur). Les résultats ont montré qu'il fallait augmenter la proportion de grands copeaux (en longueur) plus larges que 4 mm, pour avoir une meilleure combinaison des propriétés étudiées (la résistance au déchirement, l'éclatement et l'énergie à la rupture TEA) d'une pâte au bisulfite à très haut rendement. Cette étude a provoqué l'intérêt de l'usine impliquée, sur l'extension de la méthode aux autres parties du procédé de fabrication du papier, tel que présenté par Lupien et al (2001) qui exposent l'utilisation des modèles PLS, à Papier Masson Ltée,



pour identifier les causes de variation de propriétés du papier produit, comme l'énergie à la rupture dans le sens machine, l'opacité et la résistance au déchirement dans le sens travers. L'objectif de l'étude était d'identifier les variables les plus influentes sur la qualité du papier et d'établir les conditions optimales d'opération. Les résultats ont révélé que l'opération de la machine à papier était la source de variation la plus importante. L'exercice d'optimisation a permis de confirmer que l'équipement était à sa limite de capacité car les valeurs cible étaient très semblables aux valeurs actuelles d'opération.

Skoglund et al (2002) ont utilisé des données provenant d'une machine à papier carton pour comparer la performance des modèles PLS et des réseaux neuronaux décrivant un processus en continu. Bien que pour les fins de prédictions, les réseaux neuronaux ont démontré leur supériorité, les modèles PLS ont été plus adéquats pour faire le suivi du procédé, en détectant plus sensiblement les déviations des conditions normales d'opération ainsi que les variables responsables des déviations.

Stefanov et Hoo (2003) ont utilisé une approche d'analyse multivariée hiérarchisée pour déterminer les causes du phénomène de froissure du papier. Les méthodes PLS, Tucker 3 et d'analyse de facteurs parallèles PARAFAC (Parallel Factor Analysis) ont été employées. Les résultats de l'analyse ont été validés par la connaissance antérieure du procédé ce qui a démontré la capacité de l'approche d'identifier les sections du procédé et les variables responsables du phénomène de froissure du papier.

Wold et Kettaneh-Wold (2003) ont discuté l'emploi de l'analyse multivariable pour la surveillance des procédés, la détection des défauts, la prédiction des propriétés du produit et du procédé ainsi que pour mieux comprendre les procédés de mise en pâte et de fabrication du papier. Les avantages sont clairs: la maîtrise de la stabilité du procédé, la réduction de la variabilité dans la qualité du produit, l'utilisation plus efficace des matières premières ainsi qu'une meilleure préparation pour faire face aux événements futurs.

Harrison et al (2004) ont fait l'analyse de données concernant l'opération de raffineurs pendant une période de 34 mois, avec des modèles PCA et PLS. L'objectif de l'étude était de déterminer le nombre de variables latentes nécessaires pour décrire la section de raffinage dans une usine de papier journal. Le débit de la pâte produite a été la variable dominante du modèle PCA, suivie en importance par l'effet de la saison (été vs hiver), le débit de l'agent de blanchiment, le degré de blancheur de la pâte, la fraction de fibres fines, la longueur des fibres, l'énergie spécifique appliquée et l'entrefer des plaques des raffineurs.

Finalement, Hagedorn et Orccotoma (2004) présentent un survol des méthodes PCA et PLS, une introduction au développement des modèles ainsi que la façon de les interpréter pour le cas de systèmes non dynamiques. Afin d'illustrer les concepts qui y sont mentionnés, des données concernant le procédé de mise en pâte et fabrication du papier sont traitées.

Les méthodes par projections évoluent de façon continue et leur potentiel d'application rend impossible la tâche d'énumérer toutes les possibilités dans n'importe quelle industrie. Plusieurs extensions et modifications peuvent être faites selon le type des données à traiter. Quelques-unes ont été résumées par Wold et al (2001). Dans le prochain chapitre, la théorie de base des méthodes par projections est abordée.

## **CHAPITRE 4**

### **PRINCIPES DE L'ANALYSE EN COMPOSANTES PRINCIPALES (PCA) ET LES PROJECTIONS DE STRUCTURES LATENTES (PLS)**

#### **4.1 Les méthodes par projections**

Une méthode par projections est définie comme une technique d'analyse multivariée qui permet d'obtenir une représentation condensée d'un ensemble de données de  $K$  dimensions, au moyen de lignes, plans et hyperplans dans cet espace. La méthode peut s'adapter à différents objectifs d'analyse tels que la condensation et la visualisation de l'ensemble des données, la classification multivariée et l'analyse discriminante ou la détermination quantitative des corrélations entre les variables. Les méthodes peuvent être appliquées pour n'importe quel type d'ensemble de données, complet ou incomplet, pouvant contenir des bruits, où les variables peuvent être colinéaires, où le nombre de variables est plus grand que le nombre d'observations ou vice-versa. Ceci représente un avantage très intéressant pour l'analyse de procédés industriels.

Dans ce chapitre, la théorie concernant les deux méthodes par projections les plus utilisées, PCA et PLS, est abordée. Seulement les concepts de base des aspects liés à ce projet sont traités. Plusieurs documents peuvent être consultés pour une étude plus approfondie. Joliffe (1986), Geladi et Kowalski (1986), Wold et al (1987), Geladi (1988), Höskuldsson (1988), Jackson (1991), Lindgren et al (1993), Rännar et al (1994,

1995), Tenenhaus et al (1999) et Eriksson et al (2001) sont, en la matière, d'excellentes références.

Afin de faciliter la compréhension des principes des méthodes par projections, quelques termes, fréquemment utilisés, sont d'abord définis. La théorie de la méthode PCA, qui constitue la base de l'analyse multivariée, est ensuite exposée. La méthode PLS, qui en est un cas particulier important, est traitée postérieurement. Il faut préciser que la nomenclature utilisée est en grande partie celle qui a été employée par Eriksson et al (2001) afin d'être en accord avec la majorité des textes sur ce sujet.

## 4.2 Définitions

Le point de départ des méthodes d'analyse multivariée est une banque de données. Celle-ci est définie comme une matrice  $X$  de  $N$  rangées (observations) et  $K$  colonnes (variables). Les observations sont des mesures des conditions, des attributs ou des propriétés d'un système, faites à un certain moment  $i$  dans le temps (figure 4.1).

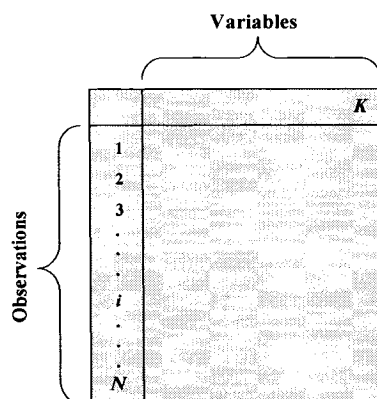


Figure 4.1 Représentation d'une matrice de données de  $K$  variables et  $N$  observations.

Lors de certains types d'analyses, il est intéressant de classer les variables selon deux catégories: les variables explicatives et les variables de réponse. Les variables explicatives sont celles dont les différents niveaux pourraient influencer la performance du système ou du procédé. Les variables de réponse sont celles qui sont mesurées afin de capturer la performance du système ou du procédé pour l'évaluer. Dans ces cas, la banque de données est constitué de deux matrices: la matrice des variables explicatives (formée par  $K$  variables et  $N$  observations) et la matrice des variables de réponse (formée par  $M$  variables et  $N$  observations), qui sont généralement appelées matrices  $X$  et  $Y$  respectivement (figure 4.2).

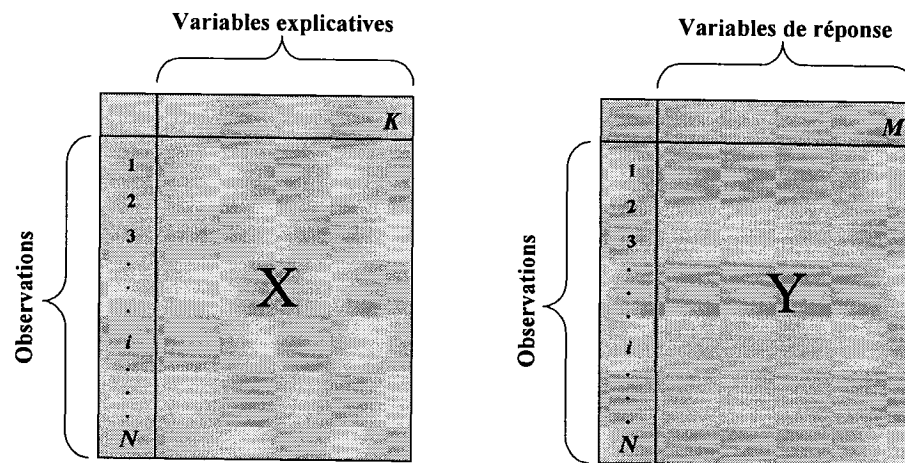


Figure 4.2 Représentation des matrices  
des variables explicatives ( $X$ ) et de réponse ( $Y$ ).

### **4.3 Prétraitement des données**

#### **4.3.1 Mise en échelle et centrage**

Avant de mettre en œuvre les méthodes PCA et PLS, on doit effectuer un traitement préalable des données afin qu'elles se conforment aux hypothèses de la méthode (Eriksson et al, 2001). Les valeurs numériques constituant une banque de données peuvent avoir des étendues très différentes. Par exemple, on peut avoir des données sur le débit d'une matière première dont les valeurs sont entre 17 000 et 30 000 L/min et sur la concentration d'une espèce chimique qui varie entre 1 et 3 ppm. Une variable dont les valeurs ont une grande étendue, comme le cas du débit de matière première de l'exemple, est caractérisée par une grande variance tandis que une variable dont les valeurs ont une petite étendue, comme le cas de la concentration de l'espèce chimique, présente une petite variance. Étant donné que PCA et PLS sont des méthodes de projection de la variance et de la covariance respectivement, les variables ayant les plus grandes variances domineront le modèle éventuellement obtenu. Si on prend les données tel qu'elles sont, le débit de matière première dominera sur la concentration de l'espèce chimique. Pour éviter ce biais artificiel, on procède à la mise à l'échelle des données.

Il y a plusieurs façons de faire la mise à l'échelle des données. La plus courante consiste à effectuer sur chaque variable la transformation linéaire ramenant sa variance à l'unité et sa moyenne à zéro. Pour chaque variable on calcule l'écart type ( $s_k$ ) dont l'inverse ( $1/s_k$ ) est le facteur de mise à l'échelle. Chaque colonne de la matrice des

données est alors multipliée par  $1/s_k$ . En outre, la valeur moyenne de chaque variable est calculée et soustraite de la valeur de chaque observation de cette variable. Toutes les variables ont alors la même variance (égale à 1) et la même moyenne (égale à zéro), et l'effet d'étendue est ainsi éliminé. Les valeurs des observations ajustées constituent un ensemble dit centré et réduit. La figure 4.3 illustre cette série d'opérations.

Une fois les variables mises à l'échelle, on peut leur attribuer des poids différents, selon l'importance qu'on voudrait donner à chaque variable dans le modèle, sur la base d'informations indépendantes disponibles sur le problème investigué.

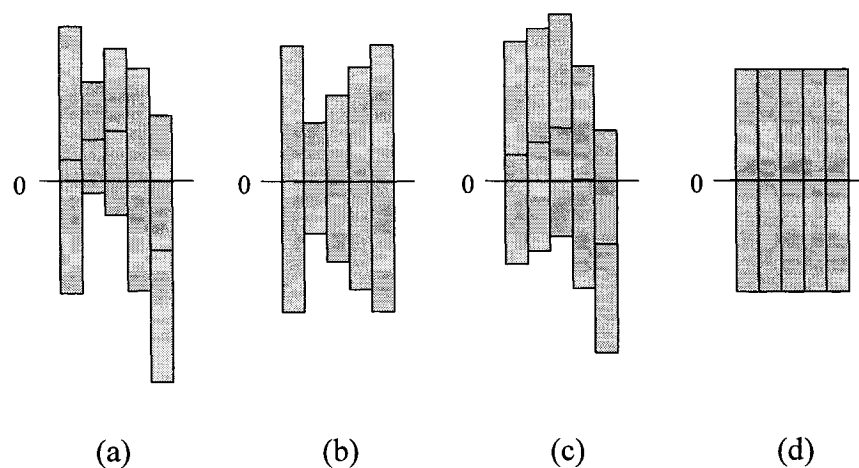


Figure 4.3 Illustration des opérations les plus courantes de mise à l'échelle. Les données concernant une variable sont représentées par une barre dont la longueur est proportionnelle à la variance des données: (a) les données brutes, (b) les données après l'opération de centrage (soustraction de la valeur moyenne), (c) les données après l'opération de réduction (division par l'écart type), (d) les données centrées-réduites.



Dans le cas où toutes les variables sont exprimées dans les mêmes unités, la mise en échelle n'est pas nécessaire. Enfin, il faut noter que l'opération de centrage peut être déconseillée pour certaines structures particulières de données (Seasholtz et Kowalski, 1992).

#### **4.3.2 Traitement des observations hors limites**

Le prétraitement des données peut faire la différence entre un modèle utile et un modèle inutile. La mise à l'échelle basée sur les écarts type et les valeurs moyennes, peut être très influencée par la présence de valeurs erronées qui s'écartent anormalement de la moyenne; ce sont les valeurs hors limites. Quand l'objectif de l'analyse est autre que l'étude des anomalies du procédé, il est recommandé d'éliminer toutes ces observations suspectes afin de construire des modèles qui reflètent de façon fiable les fluctuations réelles du procédé. La suppression des valeurs qui dépassent certaines limites préfixées est une procédure simple pour produire une estimation fiable des écarts type et des moyennes. Les données qui sont supprimées peuvent être remplacés par des cases vides dans la matrice des données, qui seront considérées comme des données manquantes, ou par une estimation raisonnable obtenue par ailleurs. On peut, par exemple, prendre la valeur de l'observation la plus proche ou la moyenne des deux valeurs les plus proches. Hoo et al (2002) donnent une attention spéciale au remplacement des valeurs hors limites.

### **4.3.3 Transformations**

Il existe d'autres techniques de prétraitement des données telles que transformations, mise en échelle avancée, correction et compression qui peuvent être nécessaires dans certains cas particuliers. Ceci est notamment le cas des données qui ont une distribution qui s'éloigne de la distribution normale ou qui serait mieux capturée par une autre technique de mise en échelle ou bien qui ont une variation systématique indésirable. Jackson (1991) et Eriksson et al (2001) peuvent être consultés à ce sujet qui dépasse le cadre de la présente étude.

### **4.3.4 Traitement des observations manquantes**

Il arrive fréquemment que l'ensemble de données disponibles est incomplet, soit à la suite de l'élimination des observations hors limites qui n'ont pas été remplacées ou en raison de la défaillance d'un appareil de mesure, des mesures hors ligne manquantes ou entachées d'erreurs, etc. Une façon d'aborder le problème est d'éliminer au complet toutes les observations pour lesquelles au moins une variable présente une valeur manquante. Cette opération peut être satisfaisante quand il y a peu de données manquantes par rapport à la taille de l'échantillon mais dans le cas contraire, on risque de perdre ainsi de l'information précieuse. Les différents algorithmes des méthodes PCA et PLS proposent plusieurs techniques alternatives à l'élimination pure et simple. Dans l'algorithme NIPALS (Nonlinear Iterative Partial Least Squares), par exemple, les composantes sont calculées par régression linéaire sur les matrices  $X$  et  $Y$  sans tenir compte de données manquantes (Wold et al, 1996 et Nelson et al, 1996). Rönner et al

(1995) traitent le problème de données manquantes pour l'algorithme *kernel* en faisant une estimation préliminaire de ces données (la moyenne de la colonne respective par exemple); par la suite les composantes sont calculées et les valeurs qui ont été estimées sont remplacées par des nouvelles estimations dérivées du modèle obtenu. Le processus est répété jusqu'à ce qu'un certain critère de convergence soit satisfait. Ceci peut être très exigeant en ressources informatiques. Ce désavantage est relevé par Dayal et MacGregor (1997a, 1997b) qui proposent des améliorations à l'algorithme *kernel* pour réduire le temps de calcul et en conséquence l'effort computationnel. D'autres possibilités sont exposées par Joliffe (1986) et plus récemment par Muteki et al (2005).

#### **4.4 L'analyse en composantes principales PCA**

L'analyse en composantes principales constitue la base de l'analyse multivariée des données. L'idée principale de la méthode PCA est de réduire le nombre de dimensions d'un ensemble de données qui contient un nombre considérable de variables interdépendantes en maintenant, autant que possible, la variation présente dans l'ensemble original. Cette réduction est réalisée par la création d'un nouvel ensemble de variables indépendantes, les composantes principales (CP). Les premières CP obtenues sont celles qui rendent compte de la plus grande partie de la variation présente dans l'ensemble original contenant toutes les variables.

Le point de départ de la méthode PCA est une matrice d'observations appelée  $X$ , qui comprend  $N$  rangées (observations) et  $K$  colonnes (variables), comme celle de la figure

4.1. La méthode consiste à trouver des lignes, des plans ou des hyperplans de l'espace à  $K$  dimensions qui s'ajustent, aussi bien que possible aux observations dans le sens des moindres carrés. Ainsi, la variance des projections des coordonnées de la ligne ou du plan est maximisée (figure 4.4).

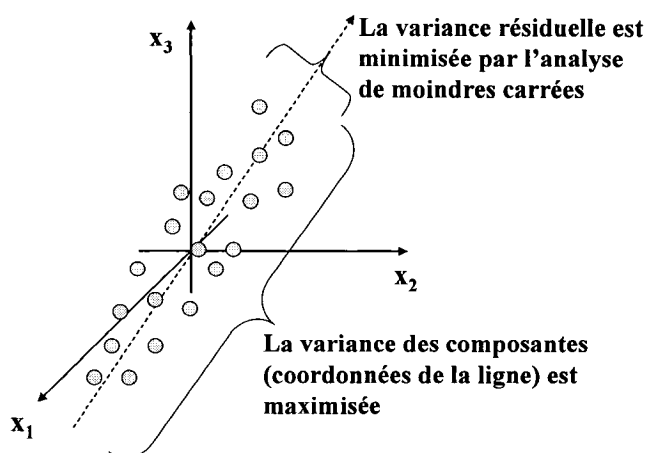


Figure 4.4 Représentation graphique d'une composante PCA (Eriksson et al, 2001).

#### 4.4.1 Interprétation géométrique de PCA

Considérons la matrice  $X$  centrée et réduite, donc dans laquelle toutes les variables ont la même étendue et leur valeur moyenne est égale à zéro. La première composante ( $a = 1$ ) est la première ligne qui s'ajuste le mieux aux observations dans le sens des moindres carrés. Cette ligne passe par la valeur moyenne (tel que montré à la figure 4.4). Chaque observation ( $i$ ) est projetée sur cette ligne pour ainsi obtenir des nouvelles coordonnées  $t_{1i}$  sur la composante.

Les valeurs  $t_{ai}$  sont définis comme la distance entre l'origine et la projection de la  $i^{\text{ème}}$  observation sur la composante  $a$  (Figure 4.5 a). Chaque composante est déterminée également par une série de poids  $p_{ak}$ . Chaque poids est défini comme le cosinus de l'angle entre la composante principale et l'axe de la variable originale (Figure 4.5 b). Ainsi, l'orientation des composantes dans l'espace à  $K$  dimensions est déterminée par les poids  $p_{ak}$ .

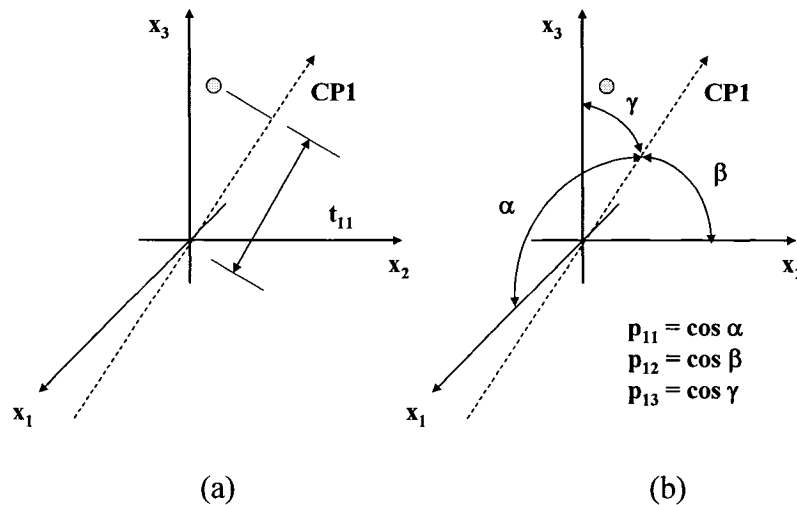


Figure 4.5 Représentation géométrique d'une composante pour  $K = 3$ .

(a) Valeurs  $t_{ai}$ . (b) Poids  $p_{ak}$ .

Dans la plupart des cas, une seule composante ne suffit pas pour capturer la totalité de la variation de l'ensemble des données. Une deuxième composante est introduite à cet effet; c'est une ligne de l'espace à  $K$  dimensions qui est orthogonale à la première composante et qui passe aussi par la valeur moyenne de l'ensemble de données en

améliorant l'approximation de la matrice  $X$ . Ces deux composantes définissent un *plan* dans l'espace à  $K$  dimensions, comme illustré à la figure 4.6 pour  $K$  égal à 3. En général, deux composantes forment un plan et plus de deux composantes forment un *hyperplan*.

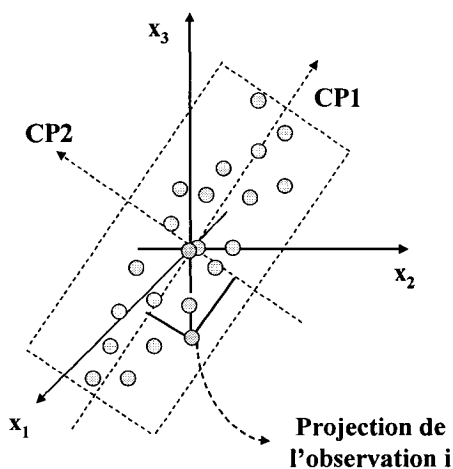


Figure 4.6 Visualisation graphique des deux composantes principales formant un plan pour  $K$  égal à 3.

La qualité descriptive du modèle PCA est améliorée en ajoutant plus de composantes. Il y a plusieurs démarches pour déterminer le nombre approprié de composantes (Jackson, 1991). Ce sujet est abordé dans la section 4.4.3. Chaque composante calculée garde la propriété d'orthogonalité par rapport aux autres composantes et passe par l'origine (la valeur moyenne).

Avec les projections de toutes les observations sur les plans définis par les composantes, il est possible de tracer un graphique de ces nouvelles coordonnées afin d'investiguer la structure de l'ensemble de données. Pour faciliter l'analyse de toutes les composantes, les plans sont retracés en prenant deux composantes à la fois. Deux types de graphiques peuvent être construits, le graphique des composantes et le graphique des poids.

Le graphique des composantes est construit avec les valeurs  $t_{ai}$ . Ce graphique montre les relations entre les regroupements des observations tandis que le graphique des poids  $p_{ak}$  montre la structure de corrélation entre les variables et permet d'identifier celles qui sont responsables des tendances décelées dans les observations. Par l'analyse conjointe des deux graphiques, on peut identifier les variables qui influencent la formation des différents regroupements d'observations. Sur le graphique des poids, les variables qui sont voisines et dans le même quadrant, sont positivement corrélées tandis que celles que se trouvent dans des quadrants diagonalement opposés, sont négativement corrélées.

D'autre part, les variables qui sont le plus éloignées de l'origine sont celles qui ont l'influence la plus significative dans le modèle. Quand il y a plus de deux composantes, toutes les combinaisons de composantes et poids doivent être considérées pour analyser le système de façon rigoureuse.

#### 4.4.2 Interprétation algébrique de PCA

Du point de vue algébrique, la méthode PCA décompose la matrice  $X$  en une série de matrices de rang 1, qui sont composées par le produit vectoriel d'un vecteur de composantes  $t_a$  et un vecteur de poids  $p_a$ :

$$X = M_1 + M_2 + \dots + M_r \quad (4.1)$$

$$X = t_1 p_1' + t_2 p_2' + \dots + t_r p_r' \quad (4.2)$$

La figure 4.7 est une représentation graphique de ces opérations. Le paramètre  $r$  est le rang de la matrice  $X$  et représente la véritable dimensionnalité fondamentale de  $X$ , c'est-à-dire, le nombre maximal de colonnes linéairement indépendantes.

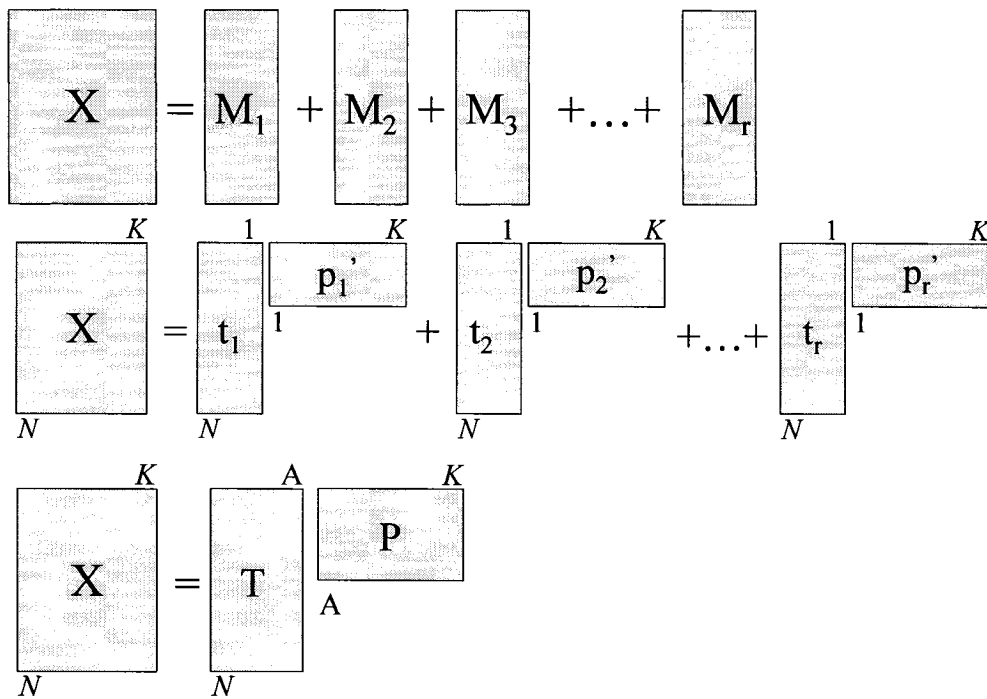


Figure 4.7 Représentation graphique de la décomposition de la matrice  $X$  en composantes.



Ainsi, le produit d'un vecteur de composantes  $t_a$  par un vecteur de poids  $p_a$  constitue une composante principale et est une combinaison linéaire des variables de l'ensemble original. Les valeurs  $t_{ai}$  sont les nouvelles coordonnées des variables latentes tandis que les poids  $p_{ak}$  représentent l'influence de chaque variable dans chaque composante.

Le modèle PCA peut être présenté comme suit,

$$X = 1 * \bar{x}' + T * P' + E \quad (4.3)$$

Le premier terme ( $1 * \bar{x}'$ ) représente les moyennes des variables calculées à l'étape de prétraitement des données. Le 2<sup>ème</sup> terme ( $T * P'$ ) modélise la structure de l'ensemble de données et le 3<sup>ème</sup> terme ( $E$ ) est la matrice des résidus (de  $K$  rangées par  $N$  colonnes), qui contient le bruit. Graphiquement:

$$\begin{array}{c} K \\ \boxed{X} \\ N \end{array} = \begin{array}{c} 1 \\ \boxed{\bar{x}'} \\ N \end{array} + \begin{array}{c} A \\ \boxed{T} \\ N \end{array} \begin{array}{c} K \\ \boxed{P'} \\ A \end{array} + \begin{array}{c} K \\ \boxed{E} \\ N \end{array}$$

Figure 4.8 Représentation graphique d'un modèle PCA.

Il y a plusieurs algorithmes pour calculer les composantes principales. La démarche classique, référée souvent comme l'algorithme *kernel*, est basée sur la théorie des valeurs et vecteurs propres (Hagedorn et Orcotoma, 2004) décrite, par exemple, dans

Strang (1980). Avec l'algorithme *kernel* PCA (Schölkopf et Smola, 2001; Lu et al, 2003), toutes les composantes principales possibles sont extraites en même temps, tandis que l'algorithme NIPALS (Geladi et Kowalski, 1986; Miyashita et al, 1990) calcule une composante à la fois, ce qui permet de mieux appréhender la méthode PCA. L'algorithme NIPALS est résumé au tableau 4.1.

Tableau 4.1 Algorithme NIPALS PCA (Geladi et Kowalski, 1986).

1. Prendre un vecteur (une colonne) $x_k$ de la matrice $X$ ; le renommer: $t_a, t_a = x_k$
2. Calculer les poids de la matrice $X, p'_a$ : $p'_a = t'_a X / t'_a t_a$
3. Normaliser les poids obtenus à la longueur 1: $p'_{a-nouveau} = p'_a / \ p'_a\ $
4. Calculer les nouvelles valeurs de la composante $a$ : $t_a = X * p_a / p'_a p_a$
5. Comparer ces nouvelles valeurs de $t_a$ à celles utilisées dans le pas 2. Si elles sont les mêmes (différence $\leq 1*10^{-10}$ par exemple), l'itération a convergé. Si elles sont différentes, aller au pas 2.
6. Calculer les résidus $E_a$ : $E_a = E_{a-1} - t_a p'_a$
7. Remplacer $X$ par $E_a$ et répéter la séquence des pas 2 à 7 le nombre de fois nécessaires pour obtenir toutes les composantes à inclure dans le modèle selon le critère choisi (voir section 4.3.3).

#### 4.4.3 Détermination du nombre de composantes principales dans un modèle PCA

Différentes méthodes ont été proposées pour déterminer le nombre de composantes qui doivent être incluses dans le modèle. Quelques exemples parmi d'autres sont l'autocorrélation, le pourcentage cumulatif de la variance, le test Scree, les moyennes des valeurs propres, la validation croisée (Li et al, 2000). La validation croisée (VC) est une méthode pratique et fiable pour tester la signifiante d'un modèle de CP et elle est devenue une approche standard en analyse multivariée (Eriksson et al, 2001). On trouve la description de cette méthode dans plusieurs documents (Wold, 1978; Eastment et Krzanowski, 1982; Geladi et Kowalski, 1986; Eriksson et al, 2001).

L'idée directrice de la VC est de maintenir une portion des données hors du modèle, de construire le modèle avec les données restantes, de prédire les données qui ont été omises par le modèle développé et de comparer les valeurs prédites avec les valeurs réelles. Ainsi, la somme des carrés des différences entre les valeurs réelles et les valeurs prédites, *PRESS* (predictive residual sum of squares), est une mesure du pouvoir prédictif du modèle. Dans SIMCA-P, le logiciel qui a été utilisé pour développer les modèles dans le présent travail, la VC est effectuée consécutivement après l'obtention de chaque composante. Pour chaque dimension, l'ensemble de données est divisé par défaut en 7 groupes et donc 7 modèles sont construits (en laissant un groupe à la fois hors du processus). Après chaque itération une *PRESS* partielle est calculée. La *PRESS* globale est la somme des *PRESS* partielles:

$$PRESS = \sum_{h=1}^{no. groups} \sum_{k=1}^K \sum_{l=1}^L (x_{hkl} - \hat{x}_{hkl})^2 \quad (4.4)$$

Dans cette équation  $x_{hkl}$  est la valeur de la variable  $k$  au temps  $l$  et  $\hat{x}_{hkl}$  est la prédiction de la valeur de la variable  $k$  au temps  $l$  pour l'itération  $h$  (Eriksson et al, 2001). À chaque nouvelle dimension, la  $PRESS$  globale est calculée et comparée avec la somme des carrés des résidus,  $RSS$  (residual sum of squares), de la dimension antérieure,

$$RSS = \sum_{k=1}^K \sum_{i=1}^I E_{ki} = \sum_{k=1}^K \sum_{i=1}^I (X_{ki} - \hat{X}_{ki})^2 \quad (4.5)$$

Ici,  $E$  désigne la matrice des résidus du modèle, et  $\hat{X}$  est la matrice des valeurs prédites par le modèle. Si la  $PRESS$  n'est pas significativement plus petite que la  $RSS$ , la composante testée est considérée comme non significative et le processus de modélisation est arrêté. La détermination du nombre de composantes à inclure dans le modèle est un compromis entre le degré d'ajustement des variables et la capacité prédictive du modèle (Eriksson et al, 2001).

Le degré d'ajustement reflète la capacité du modèle à reproduire mathématiquement l'ensemble des données. Le degré d'ajustement est quantitativement exprimé par le paramètre  $R^2$ , indice du taux de variation des données qui peut être expliquée par le modèle,

$$R^2 = 1 - RSS / SSX \quad (4.6)$$

$$SSX = \sum_{k=1}^K \sum_{i=1}^I X_{ki}^2 \quad (4.7)$$

Le terme  $SSX$  est la somme de carrés de  $X$  (sum of squares). Le degré d'ajustement peut être calculé pour deux variables:  $R^2$  indique la fraction de la variation de la première qui est expliquée par la variation de la deuxième. Pour chaque composante,  $R^2$  indique la fraction de la variation des données qui est capturée par la composante et, pour l'ensemble de composantes,  $R^2$  est la somme cumulative des  $R^2$  des composantes.

La capacité prédictive du modèle est mesurée par le paramètre  $Q^2$  qui indique la fraction de la variation des données qui est prédite par le modèle,

$$Q^2 = 1 - PRESS / SSX \quad (4.8)$$

Il faut tenir compte du fait que le  $Q^2$  global du modèle ne correspond pas à la somme cumulative des  $Q^2$  obtenus après le calcul de chaque nouvelle composante. Le  $Q^2$  global du modèle est calculé selon l'équation (4.8) en y remplaçant le ratio  $PRESS/SSX$  par le produit des ratios  $PRESS/SSX$  pour chaque composante.

Pour que  $Q^2$  soit élevé, le paramètre  $R^2$  doit aussi être élevé. Dans la plupart de cas,  $Q^2 > 0,5$  est considéré comme bon et  $Q^2 > 0,9$  comme excellent, mais ces seuils de qualité dépendent de l'application concernée (Eriksson, 2001). Les valeurs reportées de  $R^2$  et  $Q^2$  sont pour les application papetières généralement plus basses (Lupien et al, 2001; Hagedorn et Orccotoma, 2004).

#### 4.5 Les projections de structures latentes PLS

La méthode de projection de structures latentes PLS est une extension de la méthode PCA, utilisée pour relier deux matrices,  $X$  et  $Y$  (comme celles de la figure 4.2), au moyen d'un modèle linéaire multivariable. La matrice  $X$  contient les variables explicatives, appelées aussi des facteurs, et la matrice  $Y$  contient les variables de réponse. La méthode PLS peut être définie comme une technique particulière de régression pour modéliser l'association entre  $X$  et  $Y$ . La méthode PLS fonctionne dans les cas où il n'existe pas de relations explicites expliquant les relations possibles entre les variables explicatives et de réponse; elle peut être aussi utilisée pour vérifier la validité de modèles dérivés d'une théorie.

Le fondement théorique de la méthode PLS et d'autres méthodes par projections se trouve dans la théorie de la perturbation d'un système multivarié. À l'aide de cette théorie, il est possible de montrer que les modèles par projections sont capables de décrire un ensemble de données à condition qu'il y ait une certaine similarité entre les observations. Plus grande est la similarité, meilleure sera la description du système étudié par un modèle (Eriksson et al, 2001). La méthode PLS a la capacité de modéliser plusieurs variables de réponse à la fois, ce qui peut faciliter l'analyse quand les réponses sont fortement corrélées, car le nombre des modèles à considérer est réduit.

### 4.5.1 Interprétation géométrique de PLS

Comme dans le cas de la méthode PCA, chaque observation de l'ensemble de données peut être représentée par un point dans l'espace de  $X$  et un autre point dans l'espace de  $Y$ . Ceci est illustré par la figure 4.9 pour le cas de trois variables explicatives ( $K=3$ ) et trois variables de réponse ( $M=3$ ).

L'objectif est d'obtenir une bonne description des deux nuages de points et de vérifier s'il existe une relation entre les positions des observations dans les deux espaces. La première composante PLS est une ligne dans l'espace de  $X$  et une autre ligne dans l'espace de  $Y$  (figure 4.10). Les deux lignes sont calculées de façon à produire une approximation satisfaisable des deux nuages de points et une corrélation fiable entre les points projetés sur les deux. Les projections des observations sur les lignes déterminent les valeurs de composantes  $t_{1i}$  et  $u_{1i}$  pour  $X$  et  $Y$  respectivement.

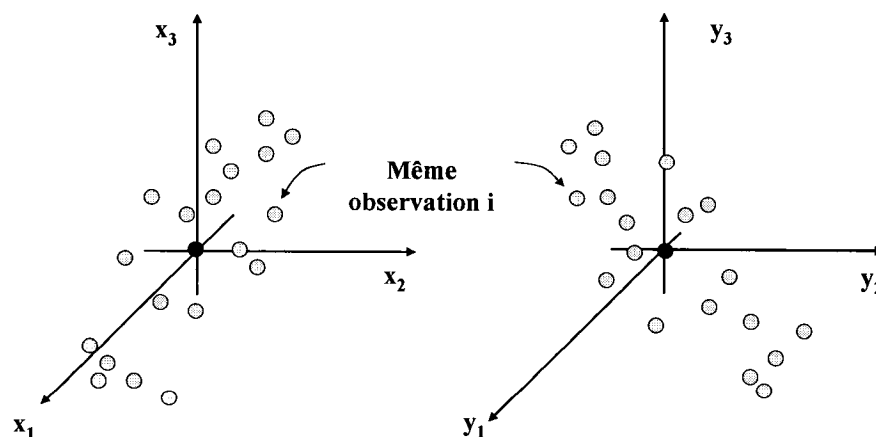


Figure 4.9 Représentation géométrique d'un ensemble des données centrées – réduites dans les espaces de  $X$  et  $Y$ .

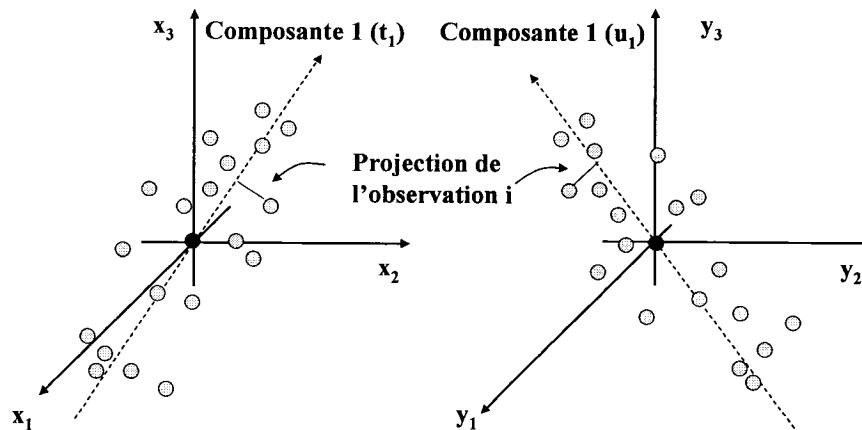


Figure 4.10 Représentation géométrique d'une composante PLS pour  $K = M = 3$ .

Les valeurs de la composante  $t_i$  et  $u_i$ , sont reliées par la relation interne  $u_{1i} = t_{1i} + h_i$ , où  $h_i$  est un résidu. Un graphique de  $t_i$  vs  $u_i$  peut être utilisé pour représenter la structure de corrélation entre  $X$  et  $Y$  et, notamment, pour mettre en évidence les points hors limites dans les espaces de  $X$  et  $Y$  ainsi que la possibilité de non-linéarité entre les variables explicatives et les variables de réponse (figure 4.11).

Une deuxième composante peut être représentée par deux autres lignes passant également par la valeur moyenne des données. Dans l'espace  $X$ , cette deuxième ligne est orthogonale à la première composante mais dans l'espace  $Y$  il n'en est pas nécessairement ainsi. L'addition d'une seconde composante permet de mieux appréhender la distribution des nuages d'observations ainsi que la corrélation entre les positions des plans dans ces espaces  $X$  et  $Y$  (figure 4.12). Géométriquement, un modèle PLS à deux composantes est constitué d'un plan dans l'espace  $X$  et d'un autre dans



l'espace  $Y$ . Les projections des observations sur ces plans donnent les valeurs des vecteurs de coordonnées  $t_a$  et  $u_a$ .

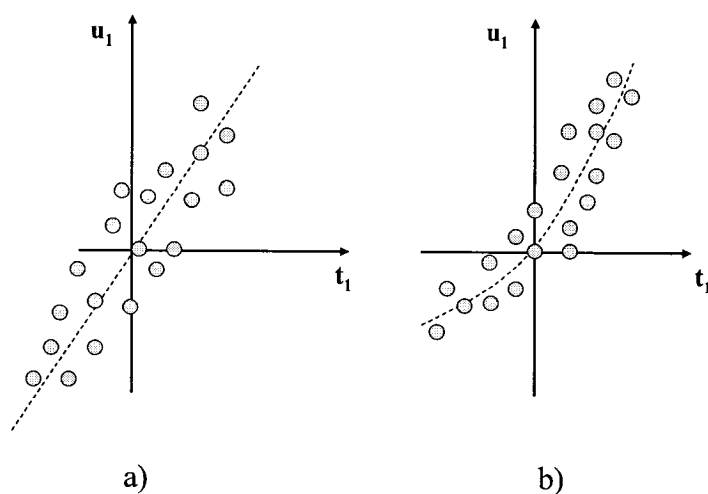


Figure 4.11 Relation interne entre les valeurs des composantes  $t_a$  et  $u_a$ .

a) Relation linéaire (la pente est égale à 1) b) Relation non-linéaire.

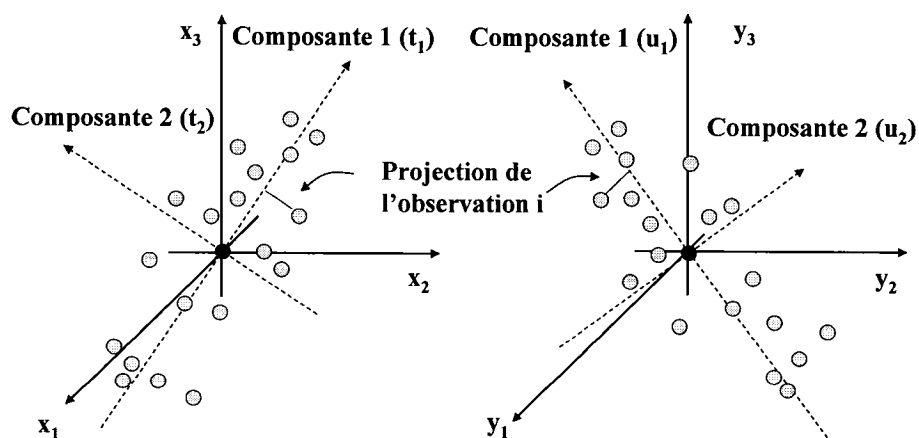


Figure 4.12 Représentation géométrique des deux composantes PLS pour  $K = M = 3$ .

Un graphique des valeurs de  $t_2$  vs  $u_2$  présente également la structure de corrélation entre  $X$  et  $Y$ . Normalement, les valeurs de la deuxième composante sont moins corrélées que les valeurs de la première étant donné que la première variable latente capture la source de la variation la plus importante (Eriksson et al, 2001). Une fois cette source de variation prise en compte par la première composante, d'autres sources de variation moins importantes sont encore présentes dans les résidus, et donc la corrélation entre  $t_2$  et  $u_2$  est évidemment plus faible que la corrélation entre  $t_1$  et  $u_1$ .

Pour interpréter un modèle PLS on doit considérer les poids des variables de  $X$  et de  $Y$ , dénotés  $w^*$  et  $c$  respectivement, qui sont interprétés de la même façon que les poids  $p_a$  d'un modèle PCA. Ces poids permettent d'identifier les variables les plus importantes dans le modèle PLS, celles qui sont regroupées, et le type de relation qui les lie. Les graphiques de valeurs des composantes et des poids sont complémentaires. La position des observations dans une direction donnée dans un graphique de composantes est influencée par les variables situées dans la même direction dans le graphique des poids.

Avec les modèles PLS on peut savoir comment obtenir une certaine réponse à partir des variables explicatives. Donc, avec la modélisation du procédé, il est possible d'en comprendre le fonctionnement mais aussi d'identifier les paramètres dont la manipulation pourrait améliorer la qualité du produit ou le rendement de la production.

#### 4.5.2 Interprétation algébrique de PLS

Comme la méthode PCA, la méthode PLS décompose l'ensemble de données en une série de matrices de rang égal à 1, qui sont formées par le produit vectoriel d'un vecteur de composantes par un vecteur de poids. La différence entre les deux méthodes est que la méthode PLS comporte deux matrices à traiter,  $X$  et  $Y$ . La matrice  $X$  est reconstituée par le produit de  $t_a$  et  $p_a$ , et la matrice  $Y$  par le produit de  $u_a$  et  $c_a$ . Le modèle PLS peut se représenter comme suit,

$$X = 1 * \bar{x}' + T * P' + E \quad (4.3)$$

$$Y = 1 * \bar{y}' + U * C' + F \quad (4.9)$$

$$u_a = b_a t_a \quad (4.10)$$

Graphiquement:

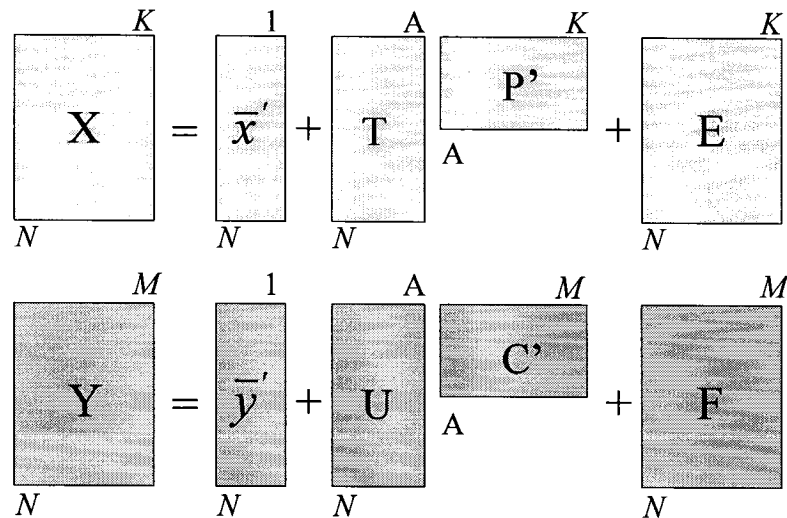


Figure 4.13 Représentation graphique d'un modèle PLS.

Le premier terme des équations (4.3) et (4.9) est le vecteur des moyennes des variables  $X$  et  $Y$ . L'information reliée aux observations est contenue dans les matrices des valeurs des composantes  $T$  et  $U$  alors que l'information concernant les variables est contenue dans les matrices de poids  $P$  et  $C$ . La variation de l'ensemble de données qui n'est pas capturée dans le modèle reste dans les matrices des résidus  $E$  et  $F$ . Les valeurs de composantes de  $X$  et  $Y$  sont reliées par l'équation (4.10). Cette relation interne est la principale différence entre la méthode PLS et la méthode par régression de composantes principales, PCR (Principal Components Regression) où les composantes principales sont calculées indépendamment pour  $X$  et  $Y$ , ce qui produit un lien plus faible entre les composantes des deux matrices. L'introduction de la relation interne dans l'algorithme de calcul des composantes force celles-ci à faire une légère rotation qui les rapproche de la ligne de corrélation (figure 4.11). Par contre, à la suite de cette opération les composantes ne sont plus orthogonales. Pour corriger cette situation les poids  $p_a$  doivent être remplacés par les poids  $w_a$  dans l'algorithme de calcul. Les poids  $w_a$  combinent les variables de  $X$  ou les résidus des variables de  $X$ , pour obtenir les valeurs des composantes  $t_a$ , en maximisant la covariance entre  $T$  et  $U$  et en conséquence entre  $T$  et  $Y$ .

Il existe un autre type de poids  $w_a^*$ , qui combinent les variables de  $X$  (et non les résidus) pour obtenir les valeurs de composantes  $t_a$ . Les matrices des poids  $w^*$  ( $W^*$ ) et  $w$  ( $W$ ), sont reliées par:

$$W^* = W(P'W)^{-1} \quad (4.11)$$

Les coefficients de régression  $B$  du modèle PLS, pour le modèle  $Y = XB$ , sont

$$B = W(P'W)^{-1}C' \quad (4.12)$$

ou bien,

$$B = W * C' \quad (4.13)$$

et

$$Y = XW * C' \quad (4.14)$$

donc,  $W^*$  est directement relié aux coefficients de régression PLS.

Les poids  $C$  caractérisent la corrélation entre les variables de  $Y$  et  $T$ . Ils combinent les variables de  $Y$  pour obtenir les valeurs des composantes  $U$  en maximisant leur corrélation avec  $X$ . Un graphique des poids  $WC$  ou  $W^*C$  montre conjointement les poids de  $X$  ( $W$  ou  $W^*$ ) et de  $Y$  ( $C$ ), et donc la structure de corrélation entre  $X$  et  $Y$ .

Finalement, il existe une autre série de valeurs qui facilite l'interprétation du modèle PLS appelées valeurs  $VIP$  (variable influence on projection). Elles mesurent l'influence de chaque variable sur la projection et représentent donc l'importance de chacune des variables de  $X$  dans le modèle. Le vecteur de valeurs  $VIP$  est une somme de carrés pondérée par les poids  $w^*$  qui tient compte de la fraction de la variance qui est expliquée par chaque dimension.

$$VIP_{Ak} = \sqrt{\left( \sum_{a=1}^A w_{ak}^2 * (SSY_{a-1} - SSY_a) \right) * \frac{K}{SSY_0 - SSY_A}} \quad (4.15)$$

Le terme  $SSY$  est la somme des carrés de  $Y$ . Pour un modèle donné, il n'y a qu'un vecteur de valeurs  $VIP$  qui condense toutes les composantes et les variables  $Y$ . Les variables explicatives qui présentent une valeur  $VIP$ , plus grande que 1, sont les plus influentes dans le modèle (Eriksson et al, 2001).

Comme pour la méthode PCA, il existe plusieurs algorithmes de calcul de PLS. Les plus communs sont les algorithmes *kernel* (Lindgren et al, 1993; De Jong et Ter Braak, 1994; Rännar et al, 1995; Dayal et McGregor, 1997; Bennett et Embrechts, 2003) et NIPALS (Geladi et Kowalski, 1986; Miyashita et al, 1990; Dayal et McGregor, 1997). Le tableau 4.2 résume l'algorithme NIPALS et permet d'avoir une meilleure compréhension de la méthode PLS.

#### 4.5.3 Détermination du nombre de composantes principales dans un modèle PLS

Il a été mentionné dans la section 4.4.3 que la validation croisée (VC) est une méthode pratique et fiable pour tester la signifiante d'un modèle de CP. Dans le cas des modèles PLS, on suit la démarche exposée plus haut; il faut néanmoins tenir compte des changements suivants dans les équations,

$$PRESS = \sum_{h=1}^{no. groups} \sum_{m=1}^M \sum_{l=1}^L (y_{hml} - \hat{y}_{hml})^2 \quad (4.16)$$

$$R^2 = 1 - RSS / SSY \quad (4.17)$$

$$SSY = \sum_{m=1}^M \sum_{l=1}^L Y_{ml}^2 \quad (4.18)$$

$$Q^2 = 1 - PRESS / SSY \quad (4.19)$$

Tableau 4.2 Algorithme NIPALS PLS (Geladi et Kowalski, 1986 avec des modifications de Dayal et MacGregor, 1997 et la notation de Eriksson et al, 2001).

1. Prendre un vecteur (une colonne)  $y_m$  de la matrice  $Y$ . Le renommer  $u_a$ :  $u_a = y_m$
2. Calculer les poids de la matrice  $X$ ,  $w'_a$ :  $w'_a = u'_a X / u'_a u_a$
3. Normaliser les poids obtenus à la longueur 1:  $w'_{a-nouveau} = w'_a / \|w'_a\|$
4. Calculer les nouvelles valeurs de la composante  $a$ :  $t_a = X \cdot w_a / w'_a w_a$
5. Calculer les poids de la matrice  $Y$ ,  $c'_a$ :  $c'_a = t'_a \cdot Y / t'_a t_a$
6. Normaliser les poids obtenus à la longueur 1:  $c'_{a-nouveau} = c'_a / \|c'_a\|$
7. Calculer les nouvelles valeurs de la composante  $a$ :  $u_a = Y \cdot c_a / c'_a c_a$
8. Comparer ces nouvelles valeurs de  $u_a$  avec celles utilisées dans le pas 2. Si elles sont les mêmes (différence  $\leq 1 \cdot 10^{-10}$  par exemple), l'itération a convergé. Si elles sont différentes, aller au pas 2.
9. Calculer les poids de la matrice  $X$ ,  $p'_a$ :  $p'_a = t'_a X / t'_a t_a$
10. Normaliser les poids obtenus à longueur 1:  $p'_{a-nouveau} = p'_a / \|p'_a\|$
11. Normaliser les nouvelles valeurs de la composante  $a$ :  $t_{a-nouveau} = t_{a-vieux} / \|p'_a\|$
12. Normaliser les poids de la matrice  $X$ ,  $w'_{a-nouveau} = w'_a / \|p'_a\|$
13. Calculer les coefficients de régression de la relation interne,  $b_a$ :  $b_a = u_a t_a / t'_a t_a$
14. Calculer les résidus de  $X$ ,  $E_a$ :  $E_a = E_{a-1} - t_a p'_a$
15. Calculer les résidus de  $Y$ ,  $F_a$ :  $F_a = F_{a-1} - b_a t_a c'_a$
16. Remplacer  $X$  par  $E_a$  et  $Y$  par  $F_a$  et répéter la séquence des pas 1 à 16 les fois nécessaires pour obtenir toutes les composantes à inclure dans le modèle selon le critère choisi (sections 4.4.3 et 4.5.3).

## **CHAPITRE 5**

### **ANALYSE DE LA VARIABILITÉ DE LA FORCE DU PAPIER DANS UNE USINE INTÉGRÉE DE PAPIER JOURNAL**

#### **5.1 Présentation<sup>2</sup>**

Le but de cette étude était d'identifier les principaux paramètres causes de variations de la force du papier qui affectent son comportement dans les presses à impression. Pour y arriver, des données mesurées en ligne, concernant les propriétés de la pâte ainsi que les paramètres d'opération des raffineurs et d'une machine à papier, ont été récupérées du système d'archivage d'une usine intégrée de papier journal. Afin d'obtenir le maximum d'information possible d'une telle base de données, une méthode d'analyse multivariée concernant l'obtention de modèles de projections de structures latentes, appelés modèles PLS multi-bloc hiérarchisés, a été utilisé.

La méthodologie suivie et les résultats obtenus sont présentés dans ce chapitre qui est autonome car il constitue une communication faite au Congrès Fibre – Papier – Carton 2005 qui sera publiée dans les comptes rendus de ce congrès. Cette communication a été faite en collaboration avec Andrea Hagedorn et Jose-Antonio Orcotoma (Paprican), Jason Baril et Bernard Bégin (Bowater Gatineau) et Jean Paris (École Polytechnique de Montréal).

---

<sup>2</sup> Cette section a pour but de faire le lien entre le présent chapitre et le mémoire.



## 5.2 Introduction

De nos jours, les usines canadiennes font face à des demandes, chaque fois plus exigeantes sur la qualité du papier. Si le produit final ne répond pas aux besoins spécifiques du client, dans la plupart des cas, la production sera vendue à un prix inférieur. D'un autre côté, les propriétés de résistance comme la tension (résistance à la rupture par traction) ou l'énergie à la rupture, TEA (Tensile Energy Absorption), sont les caractéristiques de plus en plus adoptées par les usines comme mesure de la facilité de passage sur les presses dans la salle d'impression. Une campagne de mesure antérieure (2000 - 2001) effectuée conjointement par Paprican et l'usine de Bowater Gatineau, a montré que la tension sens machine (SM) et l'énergie à la rupture SM sont les deux propriétés qui sont le mieux corrélées avec la performance du papier journal produit par cette usine dans les salles de presse. Par la suite Ferahi et Uesaka (2002) ont confirmé que la facilité de passage sur les presses est mieux prédite par la tension SM du papier que par la résistance au déchirement sens travers (ST).

Dans le but d'identifier les principaux paramètres causes de variations des propriétés mécaniques qui affectent le comportement du papier dans les presses à impression, des données mesurées en ligne, concernant les propriétés de la pâte ainsi que les paramètres d'opération des raffineurs et d'une machine à papier, ont été récupérées du système d'archivage de l'usine de Bowater Gatineau, avec laquelle ce projet a été réalisé. Des données concernant les espèces et les propriétés physiques des copeaux étaient aussi disponibles, tout comme les mesures au laboratoire des propriétés du papier produit et

des formettes faites à partir d'échantillons prélevés en plusieurs endroits du procédé de raffinage thermomécanique. La variabilité de la tension SM du papier a été choisie comme objet d'étude, étant donné sa relation de proportionnalité avec l'énergie à la rupture et la reproductibilité des mesures (5%) (TAPPI, 2000).

### **5.3 L'usine**

Cette usine intégrée de pâtes et papier journal, fabrique environ 470,000 tonnes de papier journal par année sur trois machines. Pendant la période considérée par cette étude et pour la machine étudiée, l'usine utilisait en moyenne 50% de pâte thermomécanique (PTM) et 50% de pâte desencrée (DES). L'atelier de PTM possède six raffineurs CD-76, répartis en deux lignes principales et une ligne de rejet (identifiées ici comme lignes A, B et C). La qualité de la pâte est mesurée à l'aide d'un analyseur en ligne PQM (Pulp Quality Monitor). L'atelier de DES possède trois cellules de flottation et utilise habituellement un mélange de vieux papier journal et vieux papier à magazine dans le rapport 80 à 20.

### **5.4 Données descriptives**

Le tableau 5.1 présente les variables prises en compte lors de l'analyse. Les propriétés des copeaux sont déterminées deux fois par jour, sur des composites de 12 heures. Les données sur le raffinage, le système d'approche et la machine à papier ont été récupérées du système d'archivage à intervalles d'une minute. Une première détermination de la tension est faite sur des formettes de pâte prélevée au réservoir de

mélange de PTM tous les jours à 8 h. Concernant le papier, la tension est mesurée sur chaque bobine mère produite, c'est-à-dire chaque 50 à 60 minutes en moyenne sur la machine étudiée.

## **5.5 Méthode**

### **5.5.1 Démarche de modélisation**

Dans le but de détecter les variables qui ont le plus d'influence sur la force du papier, le procédé a été divisé en deux parties, soit la mise en pâte thermomécanique et la fabrication du papier. En modélisant d'abord le procédé de mise en pâte, on identifie les variables ayant le plus d'influence sur la force des formettes. Par la suite on utilise cette information pour modéliser le procédé de fabrication de papier. On n'a pas tenu compte dans cette étude du procédé de DES, sur lequel il y a peu de moyens d'action.

Plusieurs techniques statistiques d'analyse multivariée ont été envisagées pour obtenir le maximum d'information possible d'une base de données aussi large (plus de 80 variables). La technique la plus prometteuse a été celle qui concerne l'obtention de modèles de projections de structures latentes, appelés modèles PLS multi-bloc hiérarchisés.

L'obtention de modèles PLS est une pratique courante dans l'analyse de données multidimensionnelles et colinéaires provenant de systèmes complexes. Les concepts de base, de même que divers algorithmes de calcul ont fait l'objet de travaux (Geladi et Kowalski 1996, Dayal et MacGregor 1997a, Hagedorn et Orccotoma 2004).

Tableau 5.1 Variables disponibles pour la modélisation  
de la tension SM du papier journal.

Unité du procédé	Variable
Copeaux	Espèces, densité, humidité, blancheur, classification par dimensions, éléments indésirables
Raffinage Thermomécanique	Entrefers, consistances, dilutions, âge des plaques, charges des moteurs, production, énergies spécifiques, bûchettes totales (PQM), égouttage (PQM), longueur des fibres (PQM)
Système d'approche	% cassés, % DES, % PTM, égouttage <sup>†</sup> , pH <sup>†</sup> , degré de blancheur <sup>†</sup> , agent de rétention, consist. mélange
Machine à papier	Caisse d'arrivée: pression, température, consistance, ouverture des lèvres. Pressions d'aspiration (zone de formation et des presses), consistance de l'eau blanche, turbidité, agent de blanchiment, débit du ramasse pâte, vapeur totale, vitesse du jet, vitesse de la bobine, vitesse de la toile, tirage total, humidité, épaisseur et blancheur du papier produit

<sup>†</sup> Propriétés mesurées pour la pâte mécanique, la pâte desencrée et le mélange pâte mécanique – pâte desencrée – cassés.

En dépit d'une relative simplicité, l'interprétation de résultats des modèles PLS devient compliquée lorsqu'on doit considérer un nombre important de coefficients de régression. La réduction du nombre de variables peut être une option attrayante mais on risque de perdre une partie de l'information. L'approche de mise en blocs hiérarchisés constitue une alternative intéressante. Elle consiste à regrouper les variables dans des blocs caractéristiques, à développer un modèle PLS pour chaque bloc et à utiliser les variables latentes (composantes) ainsi obtenues comme de nouvelles variables dans un

modèle PLS appelé modèle multi-bloc. Cette démarche a été présentée par Wold et al, 1996.

L'idée de base des modèles PLS multi-bloc hiérarchisés est illustrée dans ce qui suit. Dans un procédé quelconque il est possible de regrouper les variables selon leur nature ou l'opération unitaire à laquelle elles appartiennent. Pour le procédé de mise en pâte étudié (figure 5.1), on a choisi de regrouper les variables en trois blocs. D'abord les propriétés des copeaux, ensuite les variables d'opération des raffineurs et finalement les propriétés de la pâte thermomécanique produite.

Une fois le processus de mise en blocs fait, la prochaine étape consiste à obtenir un modèle PLS pour chaque bloc. Les variables latentes de chaque sous-modèle ainsi obtenues constituent ensemble le groupe de variables explicatives d'un nouveau modèle PLS, le modèle multi-bloc.

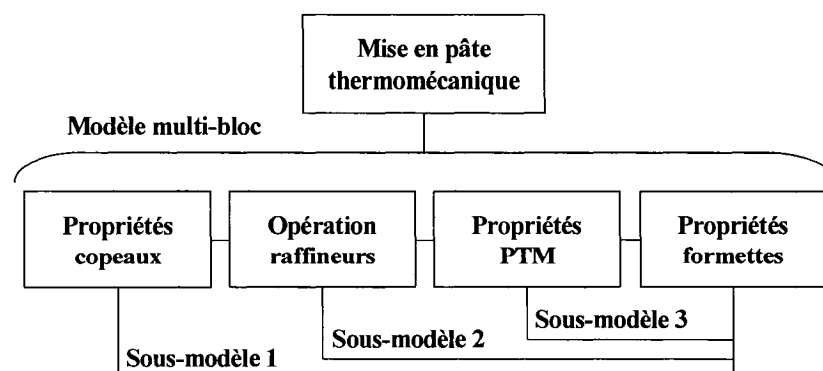


Figure 5.1. Exemple de mise en blocs et modélisation du procédé de mise en pâte.

Ainsi on peut faire la distinction entre deux niveaux. Dans le premier niveau on trouve les sous-modèles qui englobent les relations entre les variables du procédé et les propriétés des formettes et dans le deuxième niveau on trouve le modèle multi-bloc qui contient les relations entre les blocs. Il est sous-entendu que tous les modèles du premier niveau ainsi que le modèle multi-bloc doivent partager les mêmes variables de réponse. La figure 5.2 présente la façon dont on a appliqué cette démarche pour le procédé de fabrication de papier journal.

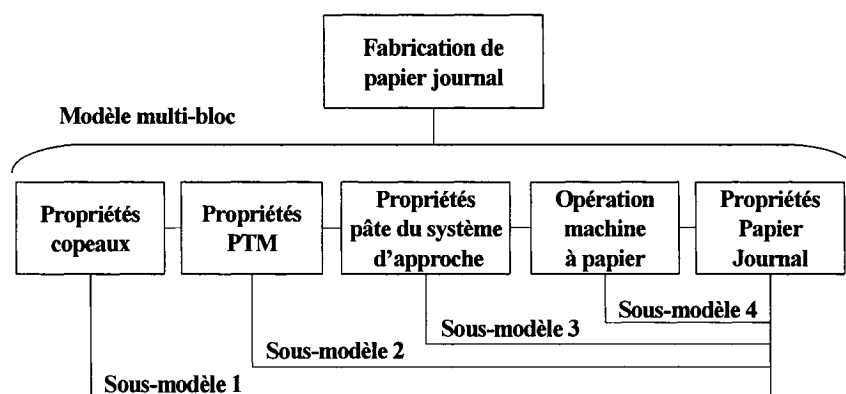


Figure 5.2. Exemple de mise en blocs et modélisation du procédé de fabrication de papier journal.

Tous les modèles ont été élaborés à l'aide du logiciel SIMCA P+, 2002. Celui-ci propose pour chaque sous-modèle, un nombre de composantes qui permet d'avoir un coefficient de détermination cumulatif  $R^2X$  (cum) minimum de 0,6 (Umetrics AB, 2002). Pour le modèle multi-bloc, le nombre de composantes significatives est déterminé par la méthode de validation croisée, basée sur le calcul de la somme des carrés des erreurs prédites (PRESS Prediction Error Sum of Squares; Eriksson et al, 2001).

## **5.5.2 Prétraitement des données**

### **5.5.2.1 Synchronisation**

Étant donné le temps de résidence de la matière dans le procédé, la synchronisation des données a été effectuée. Pour le procédé de mise en pâte, le délai estimé entre l'échantillonnage de la pâte pour faire les formettes et le raffinage des copeaux est de 10 à 12 h. Donc, on a fusionné les mesures des propriétés des copeaux échantillonnés de 20 à 8 h, avec les moyennes entre 21 et 22 h des variables d'opération du raffinage et les propriétés des formettes mesurées à 8 h le lendemain. Pour le procédé de fabrication de papier, la synchronisation n'a pas été nécessaire du fait que les délais engendrés dans le procédé étaient négligeables.

### **5.5.2.2 Élimination des observations suspectes ou hors limites**

La qualité des modèles obtenus dépend de la qualité des données, donc les observations suspectes ou hors limites, produites par exemple à cause d'une défaillance d'un appareil de mesure ou d'un arrêt de la machine à papier, ont été éliminées. Cette opération a permis de réduire la fluctuation des données.

### **5.5.2.3 Élimination de la dépendance du poids de base**

La tension du papier, tout comme l'énergie à la rupture, la résistance au déchirement, la résistance à l'éclatement, etc., sont proportionnelles au poids de base (figure 5.3(a)). Faire l'analyse des modèles avec une variable de réponse qui dépend du poids de base conduirait à une mauvaise interprétation des résultats. Afin d'éliminer une telle

dépendance, on considère l'indice de tension comme variable de réponse, c'est-à-dire la valeur de la tension divisée par le poids de base (figure 5.3(b)). L'intérêt de cette opération ne se limite pas à la production de papier à différents poids de base cibles mais est aussi avantageuse pour éliminer l'intra-variabilité dans un seul poids de base cible.

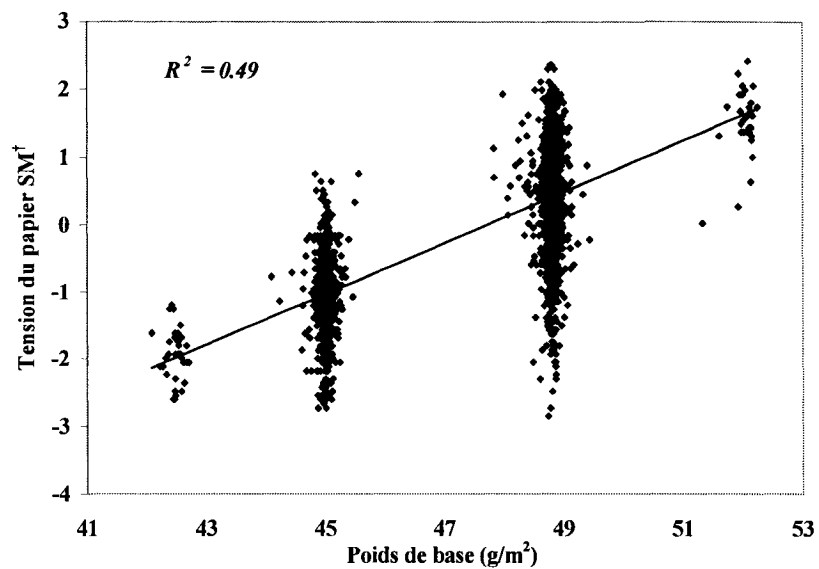
#### **5.5.2.4 Modèles pour un seul poids de base**

L'usine produit du papier journal de quatre poids de base différents. Étant donné que plus de 60% de la production de l'usine correspond au poids de base de  $48,8 \text{ g/m}^2$ , on s'est limité à l'obtention des modèles pour ce seul poids de base.

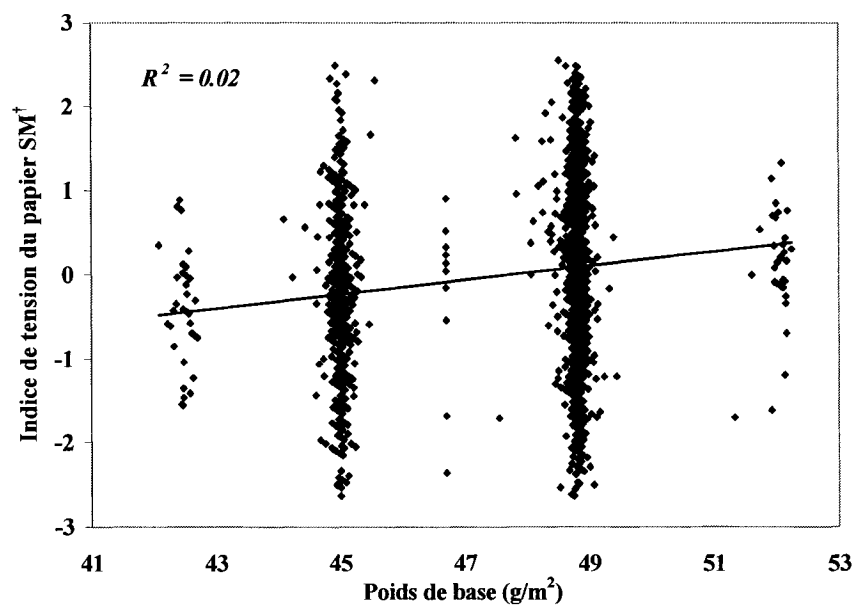
### **5.6 Résultats**

La base de données pour le poids de base de  $48,8 \text{ g/m}^2$  comprend 8 mois d'opération. Pour pouvoir tester la capacité de prédiction du modèle, les 8 mois d'opération ont été divisés en deux périodes de 4 mois. Dans la série chronologique des mesures de l'indice de tension SM du papier (figure 5.4), on peut voir que la première période présente plus de variabilité que la deuxième. Ceci est confirmé par les distributions des données et les coefficients de variation CV (l'écart type multiplié par 100 et divisé par la moyenne), montrés à la figure 5.5.





(a)



(b)

Figure 5.3. Élimination de la dépendance de la tension du papier par rapport au poids de base. (a) Le poids de base vs la tension SM du papier. (b) Le poids de base vs l'indice de tension SM du papier.  $^\dagger$ Valeurs centrées – réduites

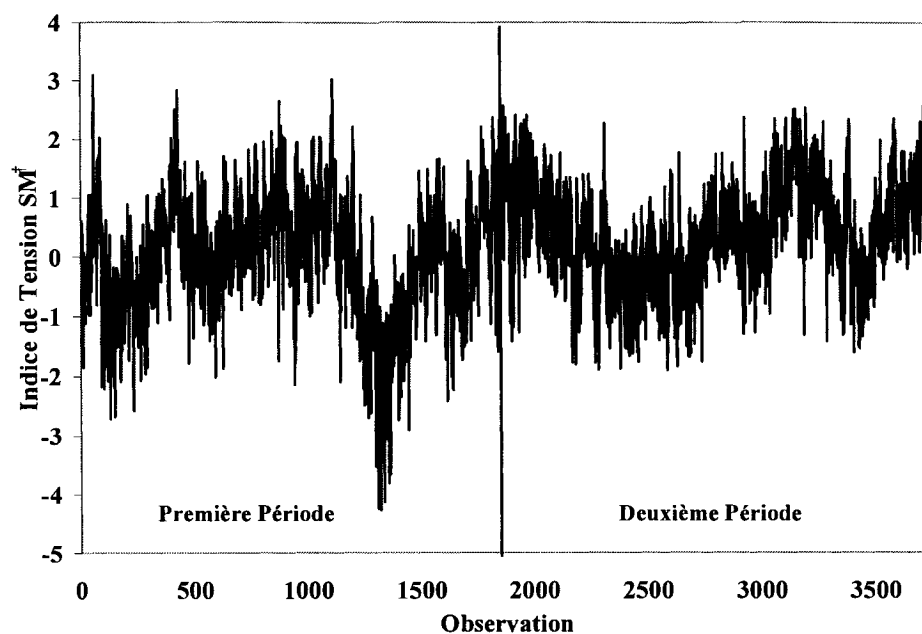


Figure 5.4. Série chronologique des mesures de l'indice de tension SM du papier.

† Valeurs centrées – réduites.

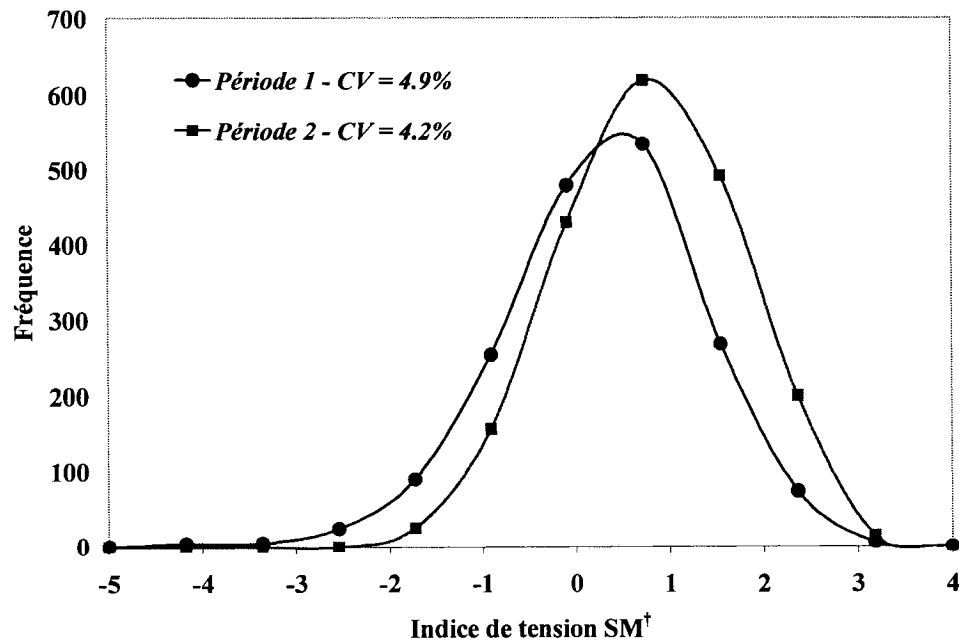


Figure 5.5. Histogramme de la base de données divisée en deux périodes.

† Valeurs centrées - réduites.

### 5.6.1 Modélisation du procédé de mise en pâte thermomécanique

Un modèle multi-bloc pour l'indice de tension des formettes, a été construit pour chaque période de 4 mois, suivant le schéma général de la figure 5.1. L'opération des raffineurs a été divisée en 5 blocs afin de traiter l'opération de chaque raffineur séparément. Les propriétés de la pâte sont aussi divisées en 3 blocs, selon la ligne de raffinage de laquelle provient la pâte. Ainsi le modèle multi-bloc est constitué des variables latentes de 9 sous-modèles.

Chaque modèle multi-bloc présente une composante significative qui rend compte de 40% de la variabilité de l'indice de tension des formettes ( $R^2Y$ ). Ceci est typique pour ce genre de système où les données sont affectées par un bruit causé par plusieurs facteurs qui influent sur la reproductibilité et la justesse des mesures (précision de la lecture, sensibilité de l'essai ou de l'appareil, erreur d'échantillonnage, mode opératoire, étalonnage de l'appareil, facteurs externes (Smook, 1997).

La figure 5.6 montre les séries chronologiques de l'indice de tension mesuré et prédit par le modèle multi-bloc pour la période respective. La prédiction du modèle suit de façon satisfaisante la réponse à long terme.

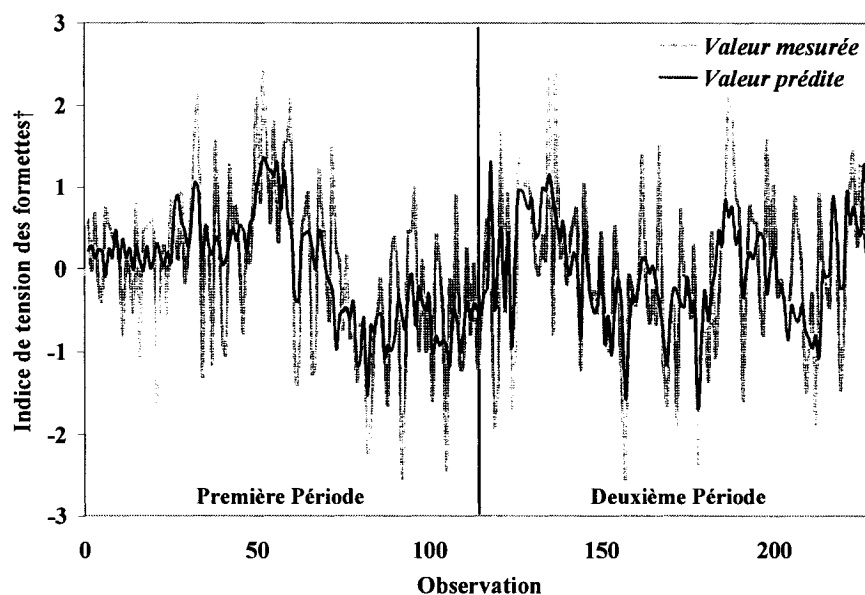


Figure 5.6. Séries chronologiques de l'indice de tension mesuré et prédit des formettes.

† Valeurs centrées - réduites.

Pour déterminer les blocs les plus influents sur la variable de réponse, on considère les valeurs VIP (Variable Importance in the Projection) de chaque variable latente. La composante avec la plus grande valeur VIP et supérieure à 1, est la plus influente sur la variable de réponse (Eriksson et al, 2001). La figure 5.7 présente en ordre décroissant, les plus grandes valeurs VIP des variables latentes pour les deux périodes. Selon ce critère, les blocs qui correspondent à la qualité de la pâte des lignes de raffinage C et B sont les plus influents sur l'indice de tension des formettes, pour chaque période respectivement.

Pour connaître les variables les plus influentes, on examine les poids  $w^*$  des variables du bloc le plus influent. Les poids  $w^*$  sont directement reliés aux coefficients de

régression PLS, ainsi les variables avec les plus grands poids (valeur absolue) sont les plus influentes sur la variable de réponse. Également, le signe du poids  $w^*$  indique le type d'influence de la variable (soit positive ou négative selon le signe).

Pour les deux périodes, la variable la plus influente concernant la qualité de la pâte des lignes C et B, est la longueur des fibres (figure 5.8). Cette observation est aussi valide pour la qualité de la pâte de la ligne A.

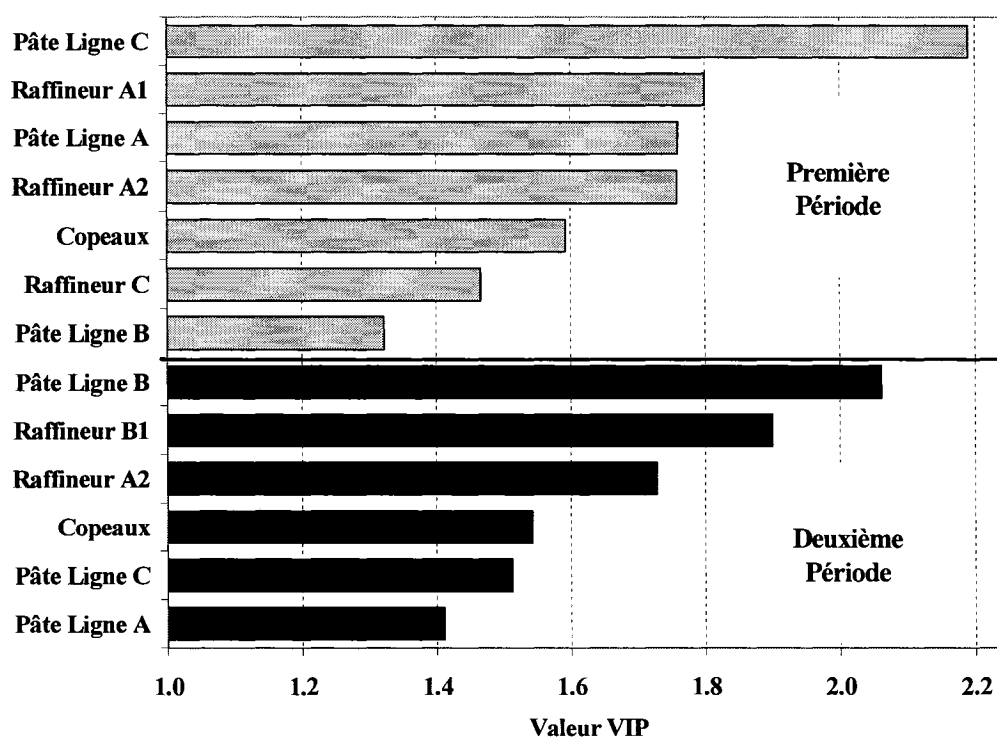


Figure 5.7. Valeurs VIP des blocs des modèles du procédé de mise en pâte.

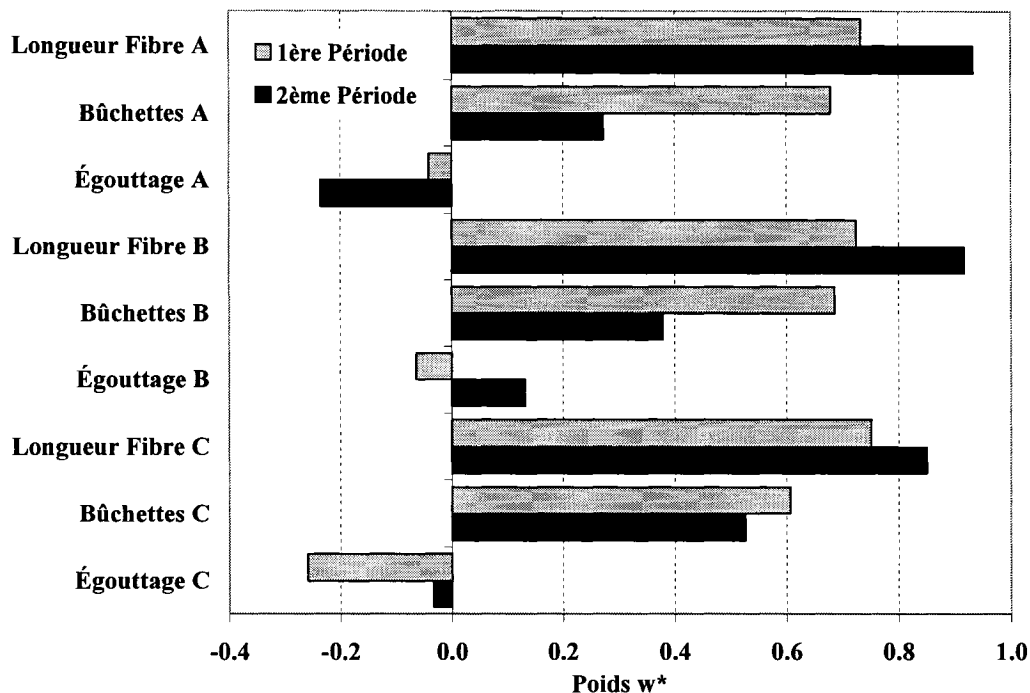


Figure 5.8. Poids  $w^*$  des variables concernant la qualité de la pâte de chaque ligne de raffinage.

On a poursuivi l'analyse en considérant les poids des variables des blocs restants concernant l'opération de raffinage. En général on a constaté que les variables les plus influentes étaient celles qui sont théoriquement connues comme étant influentes sur la longueur des fibres (l'énergie spécifique, l'entrefer, l'âge des plaques). En ce qui concerne les copeaux, les espèces ont le rôle le plus important.

D'après cette analyse, l'indice de tension des formettes augmente principalement avec la longueur des fibres tant que l'égouttage reste constant. Le coefficient de détermination  $R^2$  entre ces deux variables est de 0.22 pour la première période (pâte de

la ligne C, figure 5.9) et de 0.18 pour la deuxième période (pâte de la ligne B). Bien que les coefficients  $R^2$  peuvent sembler petits, tous les deux sont significatifs (différents de zéro) conformément au test de signifiante des coefficients de détermination (Neter, 1974). Ainsi, en ce qui concerne le procédé de mise en PTM, on peut conclure que le facteur le plus influent sur l'indice de tension des formettes est la longueur des fibres, compte tenu que l'égouttage est constant.

Une fois les blocs et les variables les plus influents sur l'indice de tension des formettes identifiés, il est possible de passer à la prochaine étape: la modélisation du procédé de fabrication du papier.

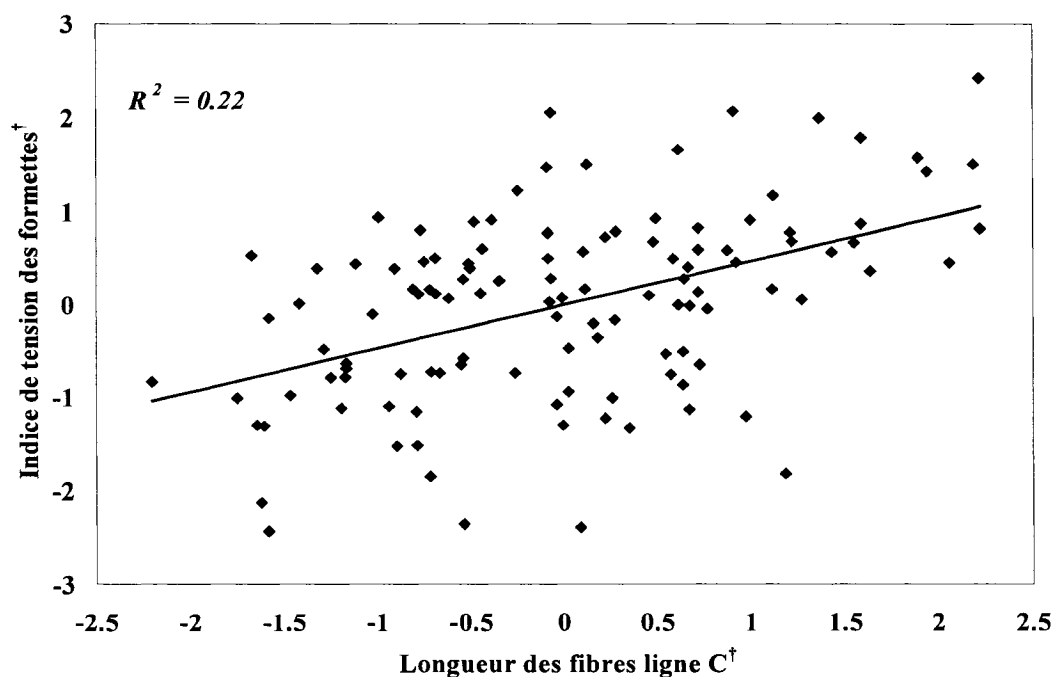


Figure 5.9. Longueur des fibres de la pâte de la ligne C vs indice de tension des formettes (1ère période). † Valeurs centrées - réduites.

### 5.6.2 Modélisation du procédé de fabrication du papier

Suivant la même approche qu'auparavant, un modèle multi-bloc pour l'indice de tension du papier à poids de base de  $48,8 \text{ g/m}^2$ , a été construit pour chaque période de 4 mois, suivant le schéma général de la figure 5.2. Conformément aux résultats précédents concernant les copeaux, seulement les données relatives aux espèces ont été prises en compte.

Le modèle multi-bloc de la première période présente trois composantes significatives qui rendent compte de 45% de la variabilité de l'indice de tension SM du papier ( $R^2Y(\text{cum})$ ), tandis que le modèle multi-bloc de la deuxième période présente deux composantes significatives qui rendent compte de jusqu'à 50% de la variabilité. La figure 5.10 montre les séries chronologiques de l'indice de tension du papier mesuré et prédit par le modèle de la période respective. La prédiction du modèle arrive à suivre la réponse à long terme, sans tenir compte du bruit de la mesure.

Le tableau 5.2 présente en ordre décroissant, les plus grandes valeurs VIP des variables latentes pour les deux périodes, et les poids  $w^*$  des variables les plus influentes de chaque bloc.

On analyse d'abord les résultats qui sont communs aux deux périodes, soit ceux qui concernent les propriétés de la pâte thermomécanique et ceux qui concernent la pâte du système d'approche.



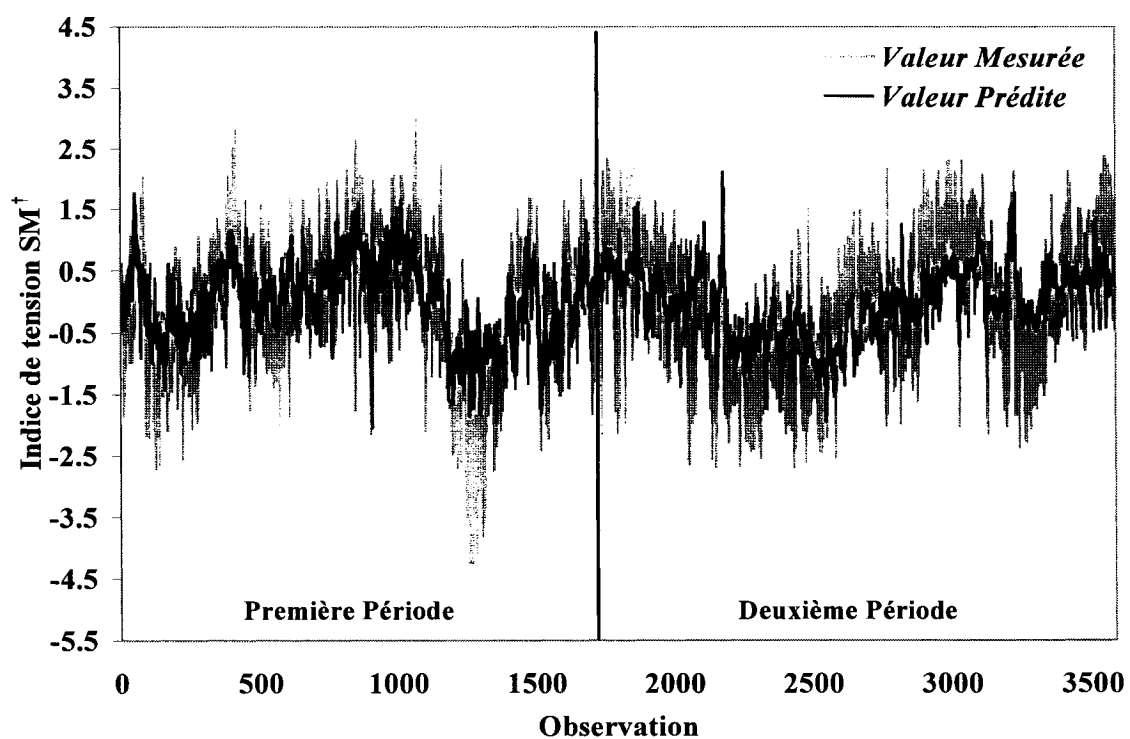


Figure 5.10. Séries chronologiques de l'indice de tension mesuré et prédit du papier.

<sup>†</sup> Valeurs centrées - réduites.

Tableau 5.2 Valeurs VIP des blocs et poids  $w^*$  des variables les plus significatives du procédé de fabrication du papier.

Première Période			
VIP	Unité de procédé	Variable	w*
2.24	Propriétés PTM	Longueur Fibre Ligne C	0.62
		Longueur Fibre Ligne A	0.57
		Longueur Fibre Ligne B	0.47
2.20	Machine à papier	Degré de blancheur	-0.53
		Débit de vapeur totale	0.45
1.55	Système d'approche	Égouttage du mélange	0.49
		% Désencrée	-0.48
		Degré blancheur PTM	0.42
		% PTM	0.32
		Débit agent de rétention	-0.27
Deuxième Période			
2.79	Machine à papier	Consistance. caisse d'arrivée	-0.40
		Épaisseur	0.34
1.97	Système d'approche	% Désencrée	-0.52
		% PTM	0.45
		Égouttage PTM	-0.38
		Degré blancheur PTM	0.24
1.56	Propriétés PTM	Longueur Fibre Ligne B	0.85
		Longueur Fibre Ligne C	0.44
		Longueur Fibre Ligne A	0.27

### 5.6.2.1 Influence des propriétés de la pâte thermomécanique

Tel que souligné par les résultats concernant le procédé de mise en pâte thermomécanique, la longueur des fibres joue un rôle très important dans le procédé de fabrication du papier. Avec un  $R^2$  entre 0.08 et 0.15, l'indice de tension SM du papier est significativement corrélé à la longueur des fibres de la pâte thermomécanique (figure 5.11).

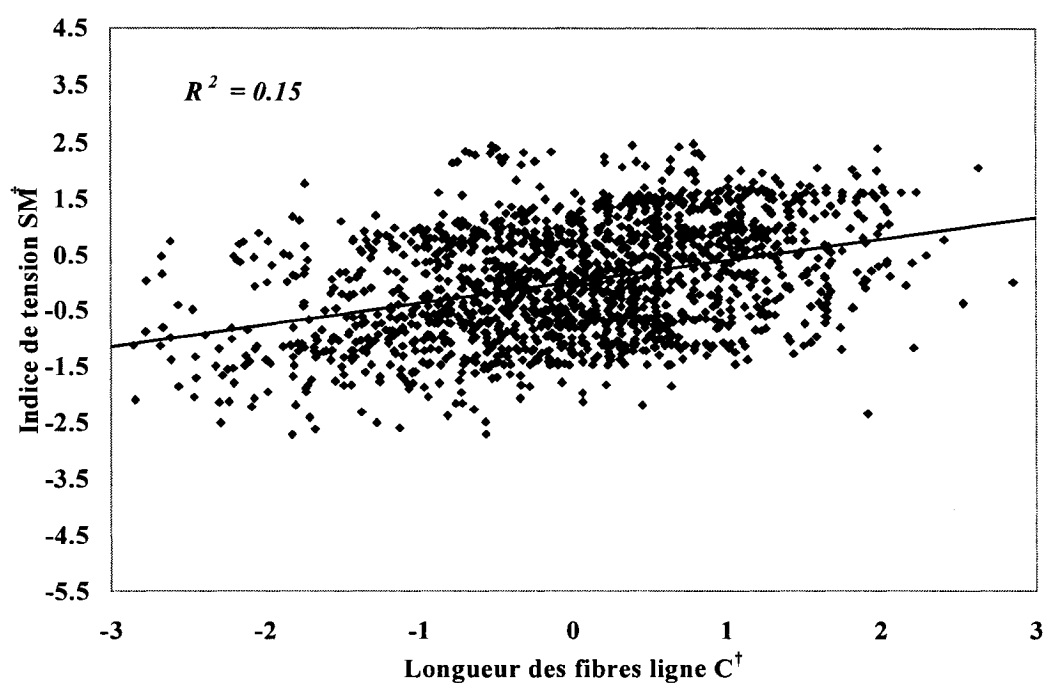


Figure 5.11. Longueur des fibres de la PTM de la ligne C vs l'indice de tension SM du papier (1ère période). <sup>†</sup> Valeurs centrées – réduites.

### 5.6.2.2 Influence des propriétés de la pâte du système d'approche

#### *Composition de la pâte*

D'après les poids  $w^*$  des variables concernant la composition de la pâte du système d'approche, la pâte DES fait diminuer l'indice de tension SM du papier, tandis que la PTM le fait monter. On ne s'attendait pas à ce résultat. On avait vu que la pâte DES est en moyenne 10% plus forte que la PTM du fait que les formettes de DES présentent un indice de tension significativement plus grand que celui des formettes de PTM (figure 5.12). Il semblerait que ceci ne se reflète pas dans le cas du papier, pour les deux périodes étudiées (figure 5.13). Ceci suggère que certains paramètres d'opération changent en fonction de la composition de la pâte et affectent négativement la tension SM du papier quand le pourcentage de pâte DES augmente. Le tirage pourrait être un de ces paramètres.

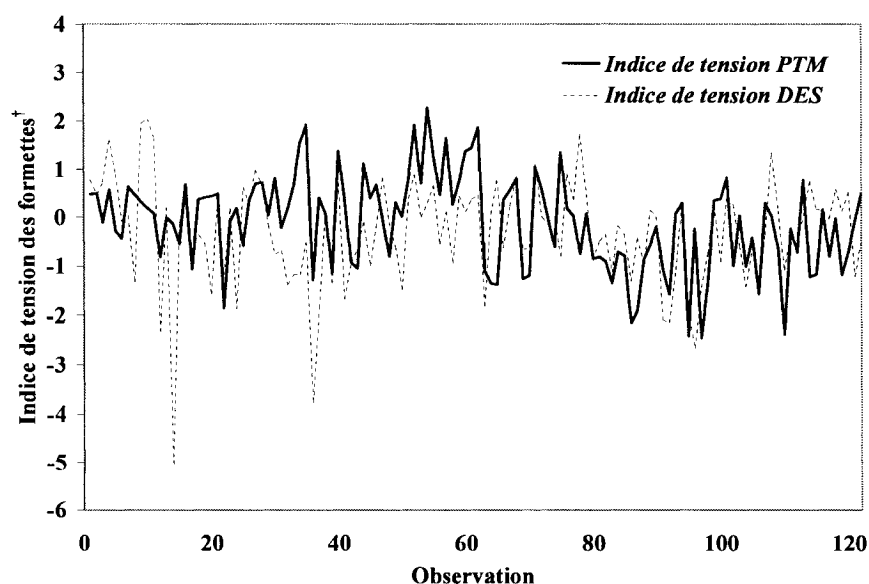
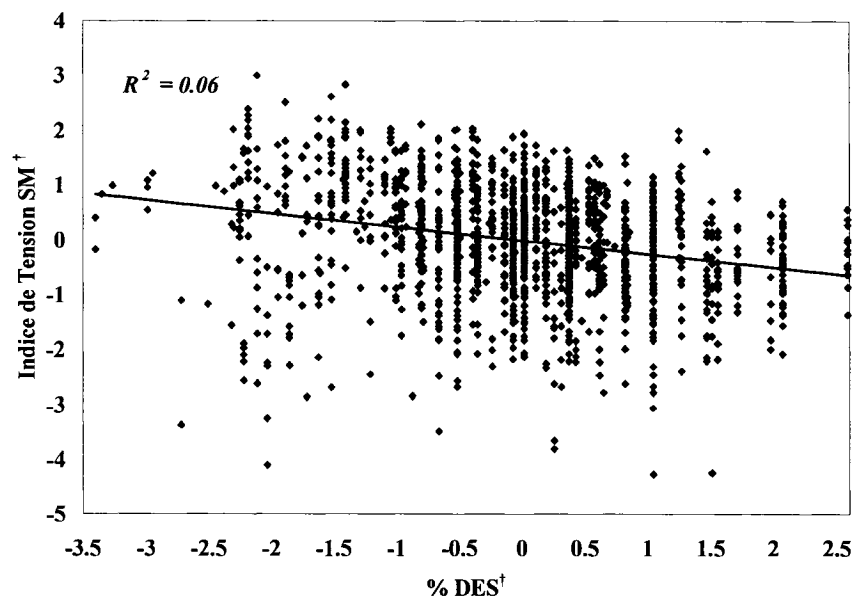
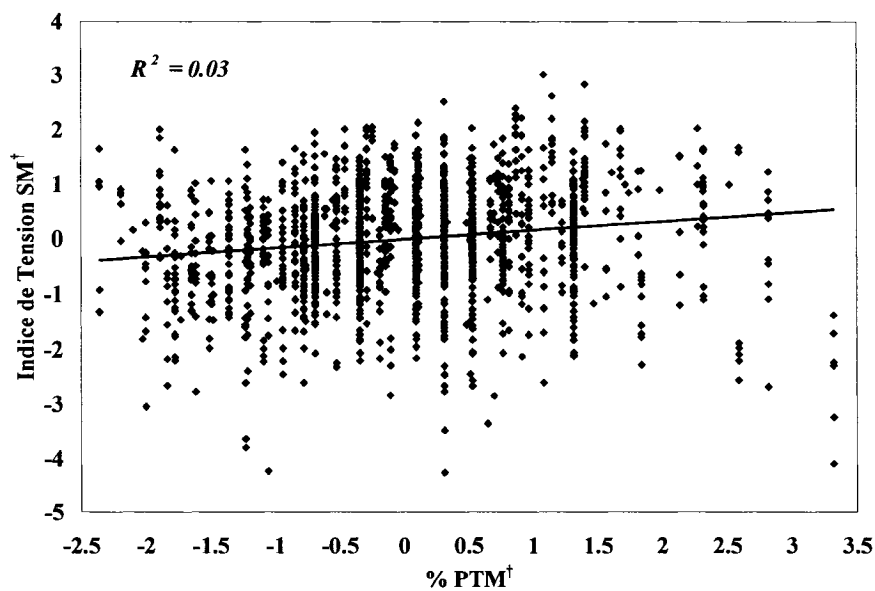


Figure 5.12. Séries chronologiques de l'indice de tension des formettes de PTM et DES (1ère période). <sup>†</sup> Valeurs centrées – réduites.



a)



b)

Figure 5.13. Influence de la composition de la pâte du circuit de tête sur la force du papier de la première période: a) %DES vs l'indice de tension SM. b) %PTM vs l'indice de tension SM. <sup>†</sup> Valeurs centrées – réduites.

Dans les espaces ouverts entre les sections de la machine on doit maintenir la feuille de papier sous tension suffisante (le tirage) pour éviter le flottement, sans toutefois amorcer de rupture. La marge d'opération est étroite. Lorsque le % DES augmente, la feuille a tendance à adhérer aux rouleaux de la machine et on doit augmenter le tirage pour maintenir la feuille dans les limites sécuritaires d'opération. D'après Pikulik (1997) l'augmentation du tirage à la section des presses pourrait affecter négativement les propriétés du papier comme le pourcentage d'élongation et l'énergie à la rupture. Par contre, dans notre travail il n'a pas été possible de vérifier que ceci s'applique à la tension SM du papier, du fait qu'il n'y avait pas des données disponibles sur le tirage ouvert de la machine et qu'on n'a pas trouvé une corrélation significative entre le tirage total et l'indice de tension SM du papier.

#### *Égouttage du mélange des pâtes et de la pâte thermomécanique*

Dans la première période, l'égouttage du mélange des pâtes thermomécanique, desencrée et de cassés apparaît comme la variable la plus influente parmi celles concernant les propriétés de la pâte du système d'approche. Pourtant l'usine réalise un très bon contrôle de ce paramètre. Les mesures fluctuent  $\pm 5\%$  du point de réglage, de la sorte qu'on ne peut pas tirer de conclusion par rapport à cette variable.

Dans la deuxième période, l'égouttage de la pâte thermomécanique apparaît comme une des variables les plus influentes. Plusieurs calibrations de la mesure pendant cette période ont affecté les données de sorte qu'elles ne sont plus fiables; on a donc écarté cette variable de l'analyse.

### *Degré de blancheur de la pâte thermomécanique*

Dans la première période, le degré de blancheur de la pâte thermomécanique apparaît comme une des variables les plus corrélées avec l'indice de tension SM. D'après les poids  $w^*$ , une pâte plus blanche donne un papier plus résistant. Examinant les causes possibles d'une telle influence, on a trouvé que le degré de blancheur de la pâte thermomécanique est significativement corrélé à la longueur des fibres de la pâte des trois lignes. Donc, le degré de blancheur n'est pas la variable qui fait directement augmenter l'indice de tension SM. Par contre il pourrait être un indicateur indirect de la force du papier.

### *Agent de rétention*

Toujours dans la première période, le débit de l'agent de rétention apparaît comme une des variables les plus influentes. L'augmentation du débit de l'agent de rétention entraîne une diminution de l'indice de tension SM. On a trouvé que la longueur des fibres est négativement corrélée au débit de l'agent de rétention d'une façon significative, ce qui conduit à poser l'hypothèse que la diminution de la longueur des fibres provoque une augmentation du débit de l'agent de rétention. Ceci entraîne l'augmentation de la rétention des fibres fines, mais aussi des autres particules provenant de la pâte recyclée et des eaux blanches qui gênent le contrôle du procédé ainsi que l'aptitude de la feuille au passage sur la machine. Il faudrait non seulement essayer de réduire l'entraînement de particules étrangères mais aussi optimiser l'opération du circuit d'eaux blanches. La minimisation du temps de résidence des

fibres fines dans la partie humide de la machine pourrait diminuer l'entraînement de vieilles fibres fines qui affectent également la qualité du papier.

### 5.6.2.3 Influence de l'opération de la machine à papier

Dans le cas de l'opération de la machine à papier étudiée, les variables qui apparaissent comme les plus significatives, ne sont pas les mêmes pour chaque période. Ceci est une indication du changement de la dynamique d'opération de la machine. Il n'est pas possible d'utiliser le modèle de la première période pour prédire l'indice de tension SM de la deuxième période (figure 5.14). Tel que montré au tableau 5.2, pour la première période, le degré de blancheur du papier produit, ainsi que le débit de vapeur totale utilisée lors du séchage du papier sont les variables les plus corrélées avec l'indice de tension SM. Pour la deuxième période, la consistance à la caisse d'arrivée et l'épaisseur du papier ont le rôle le plus important.

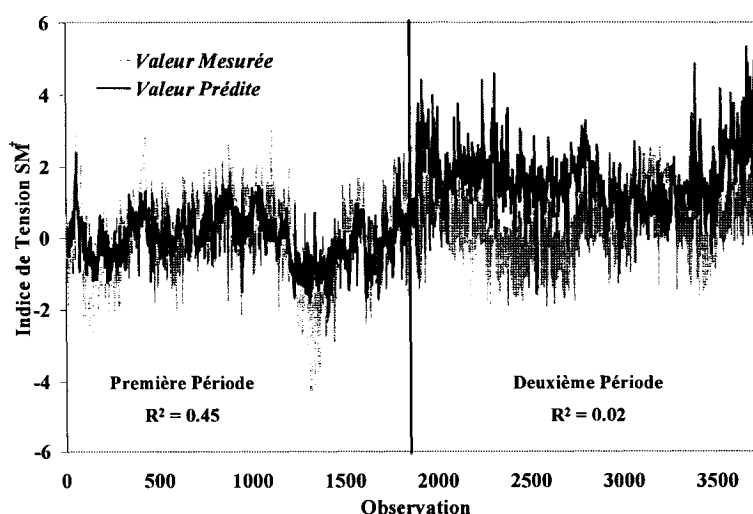


Figure 5.14. Séries chronologiques de l'indice de tension SM mesuré et prédit par le modèle de la 1ère période. <sup>†</sup> Valeurs centrées - réduites.



On ne peut pas directement attribuer l'augmentation de l'indice de tension SM du papier à l'augmentation du débit de vapeur pour le sécher. Il doit y avoir d'autres variables qui font augmenter la force du papier et par ailleurs la consommation de vapeur. Ceci s'applique aussi aux autres variables significatives qui deviennent également des variables indicatrices des changements dans le procédé qui affectent la force du papier. Afin de trouver les variables du procédé qui influencent significativement les variables d'opération de la machine, la même démarche de modélisation a été suivie. Dans ce cas, la variable de réponse était la variable d'opération étudiée et les blocs étaient formés par les variables du procédé restantes. Les résultats de cette démarche sont présentés dans ce qui suit.

#### *Degré de blancheur du papier*

L'indice de tension SM augmente avec la diminution du degré de blancheur du papier produit. Pendant la première période, il y a trois variables qui affectent significativement le degré de blancheur du papier: le pH du mélange de pâtes et le pourcentage de cassés qui le font diminuer et, le % DES qui le fait augmenter. Les deux premières variables n'ont aucun effet significatif sur la force du papier, mais le % DES présente une influence négative sur l'indice de tension SM. Ainsi, l'influence négative du degré de blancheur sur l'indice de tension SM est expliquée par la corrélation positive entre le degré de blancheur et le % DES.

Le fait que le pH est la variable la plus influente sur le degré de blancheur du papier avec, pour cette période, une influence négative, est intéressant. D'après les études antérieures faites à l'usine, le % DES en serait sans doute la cause principale.

#### *Débit de vapeur totale*

L'indice de tension SM augmente avec le débit de vapeur totale pour sécher le papier. Les variables qui affectent significativement la consommation de vapeur pendant la première période sont le pourcentage de PTM, le pourcentage de DES et l'égouttage du mélange. La consommation de vapeur augmente avec le % PTM et l'égouttage du mélange et diminue avec le pourcentage de DES. Ainsi l'influence positive du débit de vapeur sur l'indice de tension SM est expliquée par les corrélations du débit de vapeur avec la composition de la pâte du système d'approche.

#### *Consistance dans la caisse d'arrivée*

L'indice de tension SM augmente avec la diminution de la consistance de la pâte dans la caisse d'arrivée. La variable qui affecte le plus la consistance dans la caisse d'arrivée, est l'égouttage de la pâte thermomécanique. Cependant, comme on a l'a mentionné auparavant, plusieurs calibrations de la mesure effectuées pendant cette période ont conduit à écarter cette variable de l'analyse. On n'a pas trouvé de relations significatives entre la consistance dans la caisse d'arrivée et la composition de la pâte du système d'approche. Une analyse plus profonde est requise.

### *Épaisseur*

L'indice de tension SM augmente avec l'épaisseur du papier. Cependant, on a trouvé que l'épaisseur augmente aussi avec le % PTM. Ainsi, il s'agit d'un effet d'augmentation provoqué par l'épaisseur du papier, qui peut être dû à un calandrage moindre, mais aussi par la composition de la pâte du système d'approche.

## **5.7 Conclusions**

La méthode de mise en blocs hiérarchisés s'est révélée être une méthode appropriée pour l'identification des corrélations entre les différentes variables d'une large base de données, ainsi que des principales sources de variabilité dans un procédé.

La méthode a permis d'identifier et quantifier, d'une façon systématique, le rôle des étapes du procédé et des produits intermédiaires (copeaux, pâte, fibres) sur l'indice de tension sens machine du papier, dans l'usine de Bowater Gatineau:

- La longueur des fibres de la pâte thermomécanique a une influence très significative: des fibres plus longues produisent un papier plus fort en termes de tension SM.
- La composition de la pâte du système d'approche joue le rôle le plus important sur l'indice de tension SM. Une augmentation du % PTM produit un papier avec un indice de tension SM plus élevé, alors qu'une augmentation du % DES produit l'effet contraire.

- Les variables d'opération de la machine à papier qui présentent une influence significative sur l'indice de tension SM sont reliées aux effets de la composition de la pâte du système d'approche.

Il a été possible de dégager certaines mesures qui pourraient être appliquées dans la conduite du procédé pour augmenter l'indice de tension SM du papier et réduire sa variabilité:

- Optimiser la stratégie de contrôle des raffineurs afin de préserver la longueur des fibres de la PTM, tout en gardant un bon contrôle de l'égouttage.
- Augmenter le % PTM et réduire sa variabilité afin de produire un papier qui présente un bon comportement à l'impression.
- Optimiser l'opération du circuit d'eaux blanches afin de diminuer l'entraînement des particules étrangères.

L'usine a l'intention de poursuivre cette étude afin de développer un modèle prédictif de la tension SM et de l'énergie à la rupture SM du papier.

## **5.8 Remerciements**

Ce travail a pu être effectué en partie grâce à un financement du programme de recherche à la découverte du CRSNG. Monica Ortiz-Cordova, étudiante de maîtrise à l'École Polytechnique, était aussi titulaire d'une bourse du Conseil International d'Études Canadiennes.

Nous tenons aussi à remercier M. Roger Brouillette pour ses explications sur le fonctionnement du système de manutention des copeaux et le procédé de mise en pâte thermomécanique.

## 5.9 Références

DAYAL, B. S, MACGREGOR, J. F. 1997a. «Improved PLS Algorithms», *Chemometrics*, 11:1, 73 – 85.

ERIKSSON, L., JOHANSSON, E., KETTANEH-WOLD, N., WOLD, S. 2001. Multi- and Megavariate Data Analysis. Principles and Applications. . Umeå: Umetrics AB, 533 p.

FERAHI, M., UESAKA, T. 2002. *Pressroom runnability. Part II: Pressroom data analysis*. Pointe-claire, QC, Canada: Paprican P&P Report. 27 p. PPR 1502.

GELADI, P., KOWALSKI, B. R. 1996. «Partial Least-Squares Regression: A Tutorial». *Analytica Chimica Acta*. 185, p. 1 – 17.

HAGEDORN, A., ORCCOTOMA, J-A. 2004. *Multivariate data analysis: an overview of PCA and PLS*. Pointe-Claire, QC, Canada: Paprican Special Report. 30 p. PSR 530.

NETER, J., WASSERMAN W. 1974. *Applied Linear Statistical Models. Regression, Analysis of Variance and Experimental Designs*. Illinois: Richard D. Irwin, Inc. 842 p.

PIKULIK, I. I. 1997. «Wet-web properties and their effect on picking and machine runnability». *Pulp and Paper Canada*. 98:12. 161 – 165.

SIMCA-P+ version 10.0.4. 2002. Umetrics AB, Umeå.

SMOOK, G. A. 1997. *Handbook for Pulp and Paper Technologist*, 2<sup>e</sup> ed. Vancouver: Angus Wilde. 425 p.

TECHNICAL ASSOCIATION OF PULP AND PAPER INDUSTRY [TAPPI]. 2000. *Interlaboratory evaluation of test methods to determine TAPPI repeatability and reproducibility*. [Atlanta]: TAPPI Press. 58 p. EU T1200.

UMETRICS AB. 2002. SIMCA-P AND SIMCA P+ 10 Tutorial. Umeå: Umetrics AB, 130 p.

WOLD, S., KETTANEH, N., TJESSEM, K. 1996. «Hierarchical Multiblock PLS and PC Models for Easier Model Interpretation and as an Alternative to Variable Selection». *Chemometrics*. 10-5:6. 463 – 482.

### **5.10 Notes additionnelles**

Il faut préciser que les effets des variables explicatives sur les variables de réponse manifestent un de deux types de dépendance, soit une dépendance qui indique une relation de cause à effet ou une dépendance produite juste par l'existence d'une certaine corrélation. Tel qu'il a été fait dans cette étude, l'interprétation des résultats des méthodes d'analyse multivariée doit tenir compte de l'existence de ces deux types de dépendance afin d'identifier celles qui pourraient être les véritables sources de variations. On peut trouver par exemple qu'un certain facteur avec une valeur VIP élevée peut être qualifiée comme un des plus influents sur la variable de réponse. Les connaissances antérieures sur le système, des analyses de régression et de corrélation (Neter et Wasserman, 1974) ainsi que la méthodologie suivie dans cette étude (section 5.6.2.3), sont des outils qui peuvent être utilisés afin de déterminer s'il existe une véritable relation de cause à effet ou juste un effet de corrélation.

## **CHAPITRE 6**

### **CONCLUSIONS**

#### **6.1 Conclusions des résultats obtenus**

Les conclusions qui émanent des résultats obtenus, peuvent être classifiées en deux catégories: celles concernant la performance de la méthode d'analyse multivariée et celles qui touchent la conduite des procédés de mise en pâte et de fabrication du papier.

Concernant la performance de la méthode d'analyse multivariée choisie, on peut dire que la méthode de mise en blocs hiérarchisés s'est révélée être une méthode appropriée pour l'identification des corrélations entre les différentes variables d'une large base de données, ainsi que des principales sources de variabilité dans un procédé.

La méthode a permis d'identifier et de quantifier, d'une façon systématique, le rôle des étapes du procédé et des produits intermédiaires sur l'indice de tension SM du papier.

Concernant les procédés de mise en pâte et de fabrication du papier, les points saillants sont les suivants:

- La longueur des fibres de la pâte thermomécanique a une influence très significative: des fibres plus longues produisent un papier plus fort en termes de tension SM.



- La composition de la pâte du système d'approche joue le rôle le plus important sur l'indice de tension SM. Une augmentation du % PTM produit un papier avec un indice de tension SM plus élevé, alors qu'une augmentation du % DES produit l'effet contraire.
- Les variables d'opération de la machine à papier qui présentent une influence significative sur l'indice de tension SM sont reliées aux effets de la composition de la pâte fournie à la machine à papier.

## **6.2 Recommandations et perspectives**

À partir des résultats de l'analyse, il a été possible de dégager certaines mesures qui pourraient être appliquées dans la conduite du procédé pour augmenter l'indice de tension SM du papier et réduire sa variabilité:

- Optimiser la stratégie de contrôle des raffineurs afin de préserver la longueur des fibres de la PTM, tout en gardant un bon contrôle de l'égouttage.
- Augmenter le % PTM et réduire sa variabilité afin de produire un papier qui présente un bon comportement à l'impression.
- Optimiser l'opération du circuit d'eaux blanches afin de diminuer l'entraînement des particules étrangères.

L'usine a l'intention de poursuivre cette étude afin de développer un modèle prédictif de la tension SM et de l'énergie à la rupture SM du papier. Plusieurs voies peuvent être

suivies, par contre, à la lumière de la présente étude, il a été confirmé que la fabrication du papier est un processus dynamique et donc l'utilisation d'un modèle qui s'adapte aux changements dans le procédé s'impose. Le développement d'un modèle prédictif ne peut sans doute pas négliger cet aspect.

Les différentes cartes de contrôle multivariées pour faire le suivi du comportement du procédé méritent aussi d'être explorées. Leur potentiel à détecter des changements dans le procédé et leur simplicité dans la présentation des données qui facilite l'interprétation des événements, sont les avantages à exploiter.

Le développement de modèles qui englobent plus d'un grade de papier s'avère également intéressant. Non seulement il serait possible d'augmenter la connaissance du procédé avec un modèle unique qui réunit beaucoup plus d'information et qui tient compte des périodes de transition entre les grades. L'avantage à avoir un seul modèle pour faire le suivi de plusieurs grades de papier est évident, surtout pour le cas des usines qui produisent des dizaines de grades de papier.

Finalement, tel que mentionné à la section 5.5.2.1, les données utilisées dans cette étude proviennent des différentes banques de données dont la synchronisation a tenu compte des approximations des temps de résidence de la matière dans le procédé. Il serait souhaitable de réaliser une étude plus approfondie afin de vérifier la validité de ces approximations.

## RÉFÉRENCES

BARTON, R., TREDWAY, C., ELLIS, M., SULLIVAN, E. 1989. «Hydrosulfite bleaching». *Pulp and paper manufacture*. Editeur technique R.A. Leask. 3<sup>e</sup> ed. Montréal: TAPPI. P. 227 - 237.

BENNET, K. P., EMBRECHTS, M. J., 2003. «An optimization perspective on kernel partial least squares regression». *NATO Science Series III: Computer & Systems Sciences*. Editeurs J. A. K. Suykens, G. Horvath, S. Basu, C. Micchelli, J. Vandewalle. Amsterdam: IOS Press. 227 – 250.

BIERMANN, C. J. 1994. *Handbook of pulp and papermaking*. 2<sup>e</sup> ed. Toronto: Academic Press. 754 p.

CAPRON, X., WALCZAK, B., DE NOORD, O. E., MASSART, D. L. 2005. «Selection and weighting of samples in multivariate regression model updating». *Chemometrics and Intelligent Laboratory Systems*. 76:2. 205 -214.

CHEN, J., LIAO, C-M., LIN, F. R. J., LU, M-J. 2001. «Principle Component Analysis Based Control Charts with Memory Effect for Process Monitoring». *Ind. Eng. Chem. Res.* 40:6. 1516 – 1527.

CHEN, R. 1989. «Épuration de la pâte mécanique». *Pâtes mécaniques et chimico-mécaniques*. Rédacteur J. J. Garceau. Montréal: Section technique de l'Association canadienne des producteurs de pâtes et papiers. P. 175 – 199.

CHENG, S. J., MALIK, O. P., HOPE, G. S. 1988. «Damping of multi-modal oscillations in power systems using a dual-rate adaptive stabilizer». *IEEE Transactions on Power Systems*. 3:1. 101 – 108.

CHENG, S. J., MALIK, O. P., HOPE, G. S., QIN, Y. H., XU, G. Y. 1993. «An adaptive power system stabilizer based on the self-optimizing pole shifting control strategy». *IEEE Transactions on Energy Conversion*. 8:4. 639 – 645.

COOPER, J. E., WORDEN, K. 2000. «On-line physical parameter estimation with adaptive forgetting factors». *Mechanical Systems and Signal Processing*. 14:5. 705 – 730.

DAYAL, B. S., MACGREGOR, J. F., TAYLOR, P. A., MARCIKIC, S. 1994. «Application of feedforward neural networks and partial squares regression for modeling Kappa number in a continuous Kamyr digester». *Pulp and Paper Canada*. 95:1. 26 – 32.

DAYAL, B. S., MACGREGOR, J. F. 1997a. «Improved PLS Algorithms». *Journal of Chemometrics*, 11:1. 73 – 85.

DAYAL, B. S., MACGREGOR, J. F. 1997b. «Recursive exponentially weighted PLS and its applications to adaptive control and prediction». *J. Proc. Cont.* 11:1, 169 – 179.

DE JONG, S., TER BRAAK, C. J. F. 1994. «Comments on the PLS kernel algorithm». *Journal of Chemometrics*. 8:2. 169 - 174.

DENG, M. 1994. *Paper: an engineered stochastic structure*. Atlanta: TAPPI Press. 284 p.

EASTMENT, H. T., KRZANOWSKI, W. J. 1982. «Cross-Validatory Choice of the Number of Components From a Principal Component Analysis». *Technometrics*. 24:1. 73 – 77.

ERIKSSON, L., JOHANSSON, E., KETTANEH-WOLD, N., WOLD, S. 2001. *Multi- and Megavariate Data Analysis. Principles and Applications*. Umeå: Umetrics AB, 533 p.

FAHEY, M. 1989. «Testing and Control». *Pulp and paper manufacture*. Editeur technique R.A. Leask. 3<sup>e</sup> ed. Montréal: TAPPI. P. 252 – 271.

FERAHI, M., UESAKA, T. 2002. *Pressroom runnability. Part II: Pressroom data analysis*. Pointe-claire, QC, Canada: Paprican P&P Report. 27 p. PPR 1502.

FRANKO, A. 1989. «Centri-Cleaning». *Pulp and paper manufacture*. Editeur technique R.A. Leask. 3<sup>e</sup> ed. Montréal: TAPPI. P. 204 – 216.

GELADI, P. 1988. «Notes on the History and Nature of Partial Least Squares (PLS) Modelling». *Journal of Chemometrics*. 2:4. 231 – 246.

GELADI, P., KOWALSKI, B. R. 1986. «Partial Least-Squares Regression: A Tutorial». *Analytica Chimica Acta*. 185, p. 1 – 17.

HAGEDORN, A., ORCCOTOMA, J-A. 2004. *Multivariate data analysis: an overview of PCA and PLS*. Pointe-Claire, QC, Canada: Paprican Special Report. 30 p. PSR 530.

HARRISON, R. P., LEROUX, R., STUART, P.R. 2004 «Multivariate analysis of refiner operating data from a TMP newsprint mill». *Pulp and Paper Canada*. 105:4. 24 – 27.

HOO, K. A., TVARLAPATI, K. J., PIOVOSO, M. J., HAJARE, R. 2002. «A method of robust multivariate outlier replacement». *Computers and Chemical Engineering*. 26:1. 17 – 39.

HÖSKULDSSON, A. 1988. «PLS Regression Methods». *Journal of Chemometrics*. 2:3. 211 – 228.

HUNTER , J. S. 1986. «The Exponentially Weighted Moving Average» *Journal of Quality Technology*. 18:4. 203 – 210.

IVANOV I. 2001. «Multivariate techniques for process monitoring and optimization at Tembec». *Pulp and Paper Canada*. 102:7. 23 – 25.

IVANOV I., WOLD S, KETTANEH-WOLD N. 2002. «Multivariate modeling for a multigrade application». *Control Systems 2002*. Stockholm, Sweden: Pulp and Paper Technical Association of Canada. P. 180 – 183.

IVANOV, I. 2003. «Optimization of paperboard production and prediction of end-use performance using multivariate analysis». *Pulp and Paper Canada*. 104:2. 28 – 31.

JACKSON, J. E. 1980. «Principal Components and Factor Analysis: Part I – Principal Components». *Journal of Quality Technology*. 12:4. 201 – 213.

JACKSON, J. E. 1985. «Multivariate Quality Control». *Commun. Statist. – Theor. Meth.* 14:11. 2657 – 2668.

JACKSON, J. E. 1991. *A User's Guide to Principal Components*. New York: John Wiley & Sons, Inc. 569 p.

JOLLIFFE, I. T. 1986. *Principal Component Analysis*. New York: Springer-Verlag. 271 p.

KANO, M., HASEBE, A., HASHIMOTO, I., OHNO, H. 2001. «A new multivariate statistical process monitoring method using principal component analysis». *Computers and Chemical Engineering*. 25:7-8. 1103 – 1113.

KESAVAN, P., LEE, J.-H., SAUCEDO, V., KRISHNAGOPOLAN, G. A. 2000. «Partial Least Squares (PLS) based monitoring and control of batch digesters». *Journal of Process Control*. 10:2-3. 229 – 236.

KOURTI, T., NOMIKOS, P., MACGREGOR, J. F. 1995. «Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS». *Journal of Process Control*. 5:4. 277 – 284.

KOURTI, T. 2002. «Process Analysis and Abnormal Situation Detection: From Theory to Practice». *IEEE Control Systems Magazine*. 22:5. 10-25.

KRESTA, J. V., MACGREGOR, J. F., MARLIN, T. E. 1991. «Multivariate Statistical Monitoring». *The Canadian Journal of Chemical Engineering*. 69:1. 35 – 47.

LAMOUILLE, J.-L., MURRY, B., POTIÉ, C. 1989. *La Maîtrise Statistique des Procédés (SPC). Démarche et outils*. Paris: AFNOR. 126 p.

LANE, S., MARTIN, E. B., MORRIS, A. 2000. «The On-line Monitoring of Multiple Grade Production». *Control Systems 2000*, Victoria, Canada: Pulp and Paper Technical Association of Canada. P. 101 – 104.

LANE, S., MARTIN, E. B., KOOIJMANS, R., MORRIS, A. J. 2001 «Performance monitoring of a multi-product semi-batch process». *Journal of Process Control*. 11:1. 1 – 11.

LANE, S., MARTIN, E. B., MORRIS, A. 2002. «Adaptive performance monitoring of a dynamic manufacturing process». *Pulp and Paper Canada*. 103:5. 34 – 37.

LEE, Y.-H., KIM, M., CHU, Y.-H., HAN, C. 2005. «Adaptive multivariate regression modeling based on model performance assessment». *Chemometrics and Intelligent Laboratory Systems*. En cours d'impression.

LEVINSON, W. A. 1997. *SPC essentials and productivity improvement: a manufacturing approach*. Milwaukee, Wisc. : ASQ Quality Press. 266 p.

LI, W., YUE, H., VALLE-CERVANTES, S., QIN, S. J. 2000. «Recursive PCA for adaptive process monitoring». *Journal of Process Control*. 10:5. 471 – 486.

LINDGREN, F., GELADI, P., WOLD, S. 1993. «The Kernel Algorithm for PLS». *Journal of Chemometrics*. 7: 1. 45 – 59.

LOWRY, C. A., WOODALL, W. H., CHAMP, C. W., RIGDON, S. E. 1992. «A Multivariate Exponentially Weighted Moving Average Control Chart». *Technometrics*. 34:1. 46 – 53.

LOWRY, C. A., MONTGOMERY, D. C. 1995. «A review of multivariate control charts». *IIE Transactions*. 27:6. 800 – 810.

LU, C., ZHANG, T., ZHANG, R., ZHANG, C. 2003. «Adaptive Robust Kernel PCA Algorithm». *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings*. E.U.: Institute of Electrical and Electronics Engineers, Inc. [IEEE]. 6. P. 621 – 624.



LUCAS, J. M., SACCUCCI, M. S. 1990. «Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements». *Technometrics*. 32:1. 1 – 26.

LUPIEN, B., LAUZON, E., DESROCHERS, C. 2001. «PLS Modelling of strength and optical properties of newsprint at Papier Masson Ltée». *Pulp and Paper Canada*. 102:5. 19 – 21.

MACGREGOR, J. F., KOURTI, T. 1995. «Statistical Process Control of Multivariate Process». *Control Eng. Practice*. 3:3. 403 – 414.

MARTIN, E., MORRIS, J., LANE, S. 2002. «Monitoring Process Manufacturing Performance». *IEEE Control Systems Magazine*. 22:5. 26 – 39.

MIYASHITA, Y., ITOZAWA, T., KATSUMI, H., SASAKI, S. – I. 1990. «Comments on the NIPALS algorithm». *Journal of Chemometrics*. 4:1. p. 97 – 100.

MUTEKI, K., MACGREGOR, J. F., UEDA, T. 2005. «Estimation of missing data using latent variable methods with auxiliary information». *Chemometrics and Intelligent Laboratory Systems*. En cours d'impression.

NAM LAW, K. 1989. «Caractérisation des pâtes mécaniques». *Pâtes mécaniques et chimico-mécaniques*. Rédacteur J. J. Garceau. Montréal: Section technique de l'Association canadienne des producteurs de pâtes et papiers. P. 237 – 266.

NELSON, P. R. C., TAYLOR, P. A., MACGREGOR, J. F. 1996. «Missing data methods in PCA and PLS: Score calculations with incomplete observations». *Chemometrics and Intelligent Laboratory Systems*. 35:1. 45 – 65.

NETER, J., WASSERMAN W. 1974. *Applied Linear Statistical Models. Regression, Analysis of Variance and Experimental Designs*. Illinois: Richard D. Irwin, Inc. 842 p.

NOMIKOS, P. MACGREGOR, J. F. 1995. «Multivariate SPC Charts for Monitoring Batch Processes». *Technometrics*. 37:1. 41 – 59.

PIKULIK, I. I. 2001. «Wet-web properties and their effect on picking and machine runnability». *Pulp and Paper Canada*. 98:12. 161 – 165.

QIN, S. J. 1993. «Partial Least Squares Regression for Recursive System Identification». *Proceedings of the 32<sup>nd</sup> Conference on Decision and Control Proceedings*. Texas: Institute of Electrical and Electronics Engineers, Inc. [IEEE]. 3. P. 2617 – 2621.

QIN, S. J. 1998. «Recursive PLS algorithms for adaptive data modeling». *Computers chem. Engng*. 2:4-5. 503 – 514.

RÄNNAR, S., LINDGREN, F., GELADI, P., WOLD, S. 1994. «A PLS kernel algorithm for data sets with many variables and few objects. Part I: Theory and algorithm». *Journal of Chemometrics*. 8:2. 111 – 125.

RÄNNAR, S., GELADI, P., LINDGREN, F., WOLD, S. 1995. «A PLS kernel algorithm for data sets with many variables and few objects. Part II: Cross validation, missing data and examples». *Journal of Chemometrics*. 9:6. 459 – 470.

RÄNNAR, S., MACGREGOR, J. F., WOLD, S. 1998. «Adaptive batch monitoring using hierarchical PCA». *Chemometrics and Intelligent Laboratory Systems*. 41:1. 73 – 81.

SCHÖLKOPF, B., SMOLA A. J., MULLER, K.-R. 1997. «Kernel principal components analysis». *Lecture Notes on Computer Science*. Germany: Springer. 1327, p. 583 – 588.

SCRANTON, R., RUNGER, G. C., KEATS, J. B., MONTGOMERY, D. C. 1996. «Efficient shift detection using multivariate exponentially – weighted moving average control charts and principal components». *Quality and Reliability Engineering International*. 12:3. 165 – 171.

SEASHOLTZ, M. B., KOWALSKI, B. R. 1992. «The effect of mean centering on prediction in multivariate calibration». *Journal of Chemometrics*. 6:2. 103 – 111.

SIMCA-P+ version 10.0.4. 2002. Umetrics AB, Umeå.

SKOGLUND, A., BRUNDIN, A., MANDENIUS, C-F. 2002. «Comparison between linear and nonlinear prediction models for monitoring of a paperboard machine». *Chem. Eng. Technol.* 25:2. 197 – 202.

SMOOK, G. A. 1997. *Handbook for Pulp and Paper Technologist*. 2<sup>e</sup> ed. Vancouver: Angus Wilde. 425 p.

SMOOK, G. A. 2001. *Handbook for Pulp and Paper Technologist*. 3<sup>e</sup> ed. Vancouver: Angus Wilde. 425 p.

STEFANOV, Z. I., HOO, K. A. 2003. «Hierarchical multivariate analysis of cockle phenomena». *Journal of Chemometrics*. 17:10. 550 – 568.

STRANG, G. 1988. *Linear algebra and its applications*. 3<sup>e</sup> ed. San Diego: Harcourt, Brace, Jovanovich, Publishers. 505 p.

TECHNICAL ASSOCIATION OF PULP AND PAPER INDUSTRY [TAPPI]. 2000. *Interlaboratory evaluation of test methods to determine TAPPI repeatability and reproducibility*. [Atlanta]: TAPPI Press. 58 p. EU T1200.

TENENHAUS, M. 1995. «Régression PLS et applications». *Rev. Statistique Appliquée*. 43:1. 7 – 63.

TENENHAUS, M. 1999. «L'approche PLS». *Rev. Statistique Appliquée*. 47:2. P. 5 – 40.

TEPPOLA, P., MUJUNEN, S.-P., MINKKINEN, P., PUJOLA, T., PURSIHEIMO, P. 1998. «Principal components analysis, contribution plots and feature weights in the monitoring of sequential process data from a paper machine's wet end». *Chemometrics and Intelligent Laboratory Systems*. 44:1-2. 307 – 317.

TESSIER, P., BRODERICK, G., DESROCHERS, C. 1999. «Chips size distribution for an ultra-high-yield sulfite process». *TAPPI Journal*. 82:12. 1 – 7.

TSUNG, F. 1999. «Improving automatic-controlled process quality using adaptive principal component monitoring». *Qual. Reliab. Engng. Int.* 15:2. 135 – 142.

UMETRICS AB. 2002. *SIMCA-P AND SIMCA P+ 10 Tutorial*. Umeå: Umetrics AB, 130 p.

VIJAYSAI, P., GUDI, R. D., LAKSHMINARAYANAN, S. 2003. «Identification on Demand Using a Blockwise Recursive Partial Least-Squares Technique». *Ind. Eng. Chem. Res.* 42:3. 540 – 554.

WACHS, A., LEWIN, D. R. 1999. «Improved PCA Methods for Process Disturbance and Failure Identification». *AIChE Journal*. 45:8. 1688 – 1700.

WANG, X., KRUGER, U., LENNOX, B. 2003. «Recursive partial least squares algorithms for monitoring complex industrial processes». *Control Engineering Practice*. 11:6. 613 – 632.

WESTERHUIS, J.A., KOURTI, T., MACGREGOR, J. F. 1998. «Analysis of Multiblock and Hierarchical PCA and PLS Models». *Journal of Chemometrics*. 12:5. 301 – 321.

WHEELER, D. J. 1995. *Advanced topics in statistical process control: the power of Shewhart's charts*. Knoxville, TN : SPC Press. 470 p.

WISE, S. A., FAIR, D. C. 1998. *Innovative control charting: practical SPC solutions for today's manufacturing environment*. Milwaukee, Wisc. : ASQ Quality Press. 292 p.

WOLD, S. 1978. «Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models». *Technometrics*. 20:4. 397 – 404.

WOLD, S., ESBENSEN, K., GELADI, P. 1987. «Principal Components Analysis». *Chemometrics and Intelligent Laboratory Systems*. 2:1. 37 – 52.

WOLD, S. 1994. «Exponentially weighted moving principal components analysis and projections to latent structures». *Chemometrics and Intelligent Laboratory Systems*. 23:1. 149 – 161.

WOLD, S., KETTANEH, N., TJESSEM, K. 1996. «Hierarchical Multiblock PLS and PC Models for Easier Model Interpretation and as an Alternative to Variable Selection». *Journal of Chemometrics*. 10:5-6. 463 – 482.

WOLD, S., TRYGG, J., BERGLUND, A., ANTII, H. 2001. «Some recent developments in PLS modeling». *Chemometrics and Intelligent Laboratory Systems*. 58:2. 131 – 150.

WOLD, S., KETTANEH-WOLD, N. 2003. «Improving pulp and paper process diagnostics and knowledge by means of multivariate analytical techniques (MVA)». *Pulp and Paper Canada*. 104:5. 48 – 50.

ZHANG, H., TANGIRALA, A. K., SHAH, S. L. 1999. «Dynamic Process Monitoring using Multiscale PCA». *IEEE Canadian Conference on Electrical and Computer Engineering*. Edmonton: Institute of Electrical and Electronics Engineers, Inc. [IEEE]. 3. P. 1579 – 1584.

ZULLO, L. 1996. «Validation and Verification of Continuous Plants Operating Modes Using Multivariate Statistical Methods». *Computers chem. Engng.* 20, Suppl. 683 – 688.

**Ressources sur Internet**

HUBBE, M. 2004. *Mini-encyclopedia of papermaking wet-end chemistry*. [En ligne]. <http://www4.ncsu.edu/~hubbe/EquipUnit/TopFormmr.htm>. (Page consultée le 17 mai 2005)

SCHÖLKOPF, Bernhard, SMOLA, Alexander J. 2001. *Learning with Kernels. Support Vector Machines, Regularization, Optimization and Beyond*. [En ligne]. The MIT Press. <http://www.learning-with-kernels.org/> (Page consultée le 14 avril 2005)

STRATEGIS. 2005. Tableaux de données – SCIAN 322. In. Strategis: Le site canadien des entreprises et des consommateurs. [En ligne]. [http://strategis.ic.gc.ca/canadian\\_industry\\_statistics/cis.nsf/IDF/cis322datf.html](http://strategis.ic.gc.ca/canadian_industry_statistics/cis.nsf/IDF/cis322datf.html) (Page consultée le 17 mai 2005)

## ANNEXES

### A. TEST DE SIGNIFIANCE DU COEFFICIENT DE DÉTERMINATION

Test d'hypothèse du coefficient de détermination d'une population bivariée,  $\rho_{12}$  (Neter et Wasserman, 1974).

$$H_0 : \rho_{12} = 0$$

$$H_a : \rho_{12} \neq 0$$

Si  $|t^*| \leq t(1-\alpha/2; n-2)$ , on ne rejette pas  $H_0$

Si  $|t^*| > t(1-\alpha/2; n-2)$ , on ne rejette pas  $H_a$

Où  $t$  est le percentile de la distribution  $t$  de Student correspondant au niveau de confiance  $\alpha$  et  $(n-2)$  degrés de liberté ( $n$  étant le nombre d'observations).

$t^*$  est calculé selon:

$$t^* = \frac{r_{12} \sqrt{n-2}}{\sqrt{1-r_{12}^2}} \quad (\text{A.1})$$

Où  $r_{12}$  est le coefficient de corrélation de l'échantillon.

Exemple:

Avec  $n = 1651$ ,  $r_{12}^2 = 0.15$ ,  $\alpha = 10\%$

$$t^* = 17.0587$$

$t = 1.645$  (Neter et Wasserman, 1974).

$|17.0587| > 1.645$  donc on ne rejette pas  $H_a$  et on conclut que le coefficient de détermination est significativement différent de zéro.



## ANNEXE B

### COEFFICIENTS DES MODÈLES OBTENUS

B.1 Première Période (Juin à Septembre 2003).

B.1.1 Modèles ayant comme variable de réponse l'indice de tension des formettes.

Tableau B.1.1.1 Coefficients  $R^2X$ ,  $R^2Y$  et  $Q^2$  du modèle PLS multi-bloc  
et des sous modèles.

Modèle PLS Multi-bloc de l'Indice de Tension des Formettes - 1ère période						
No. d'observations (N)=117, No. de variables explicatives (K)=31						
A	$R^2X$	$R^2X(cum)$	$R^2Y$	$R^2Y(cum)$	$Q^2$	$Q^2(cum)$
1	0.1530	0.1530	0.3940	0.3940	0.3210	0.3210
Sous - modèle PLS - Propriétés des Copeaux						
K = 20						
1	0.2000	0.2000	0.1410	0.1410	0.0259	0.0259
2	0.1030	0.3030	0.0782	0.2190	-0.0923	-0.0641
3	0.1100	0.4130	0.0172	0.2360	-0.1670	-0.1700
4	0.1020	0.5150	0.0060	0.2420	-0.1820	-0.2880
5	0.0584	0.5730	0.0036	0.2450	-0.2320	-0.4160
6	0.0404	0.6140	0.0038	0.2490	-0.1960	-0.5580
Sous - modèle PLS - Raffineur Primaire Ligne A						
K = 11						
1	0.1910	0.1910	0.1780	0.1780	0.0916	0.0916
2	0.1170	0.3090	0.0283	0.2060	-0.1680	0.0008
3	0.1580	0.4670	0.0038	0.2100	-0.1540	-0.0992
4	0.1210	0.5880	0.0020	0.2120	-0.1730	-0.2090
5	0.1010	0.6890	0.0009	0.2130	-0.1830	-0.3300
Sous - modèle PLS - Raffineur Secondaire Ligne A						
K = 11						
1	0.2560	0.2560	0.1720	0.1720	0.0976	0.0976
2	0.1050	0.3610	0.0551	0.2270	-0.0887	0.0176
3	0.0912	0.4520	0.0090	0.2360	-0.1710	-0.0807
4	0.1490	0.6010	0.0020	0.2380	-0.1340	-0.1890
Sous - modèle PLS - Mesures PQM Ligne A						
K = 3						
1	0.5430	0.5430	0.1740	0.1740	0.1550	0.1550
2	0.3130	0.8560	0.0051	0.1790	-0.0118	0.1450

Tableau B.1.1.1 Coefficients  $R^2X$ ,  $R^2Y$  et  $Q^2$  du modèle PLS multi-bloc  
et des sous modèles (continuation).

A	$R^2X$	$R^2X(\text{cum})$	$R^2Y$	$R^2Y(\text{cum})$	$Q^2$	$Q^2(\text{cum})$
Sous - modèle PLS - Raffineur Primaire Ligne B						
K = 9						
1	0.1850	0.1850	0.0848	0.0848	0.0218	0.0218
2	0.2020	0.3870	0.0224	0.1070	-0.0477	-0.0249
3	0.1480	0.5350	0.0129	0.1200	-0.0675	-0.0940
4	0.0886	0.6240	0.0075	0.1280	-0.0634	-0.1630
Sous - modèle PLS - Raffineur Secondaire Ligne B						
K = 10						
1	0.3710	0.3710	0.0753	0.0753	0.0358	0.0358
2	0.1360	0.5060	0.0278	0.1030	-0.0422	-0.0049
3	0.1150	0.6220	0.0111	0.1140	-0.0583	-0.0635
Sous - modèle PLS - Mesures PQM Ligne B						
K = 3						
1	0.5540	0.5540	0.0827	0.0827	0.0620	0.0620
2	0.2690	0.8230	0.0255	0.1080	0.0149	0.0760
Sous - modèle PLS - Raffineur Ligne C						
K = 10						
1	0.2890	0.2890	0.1130	0.1130	0.0324	0.0324
2	0.2530	0.5420	0.0361	0.1490	-0.0436	-0.0098
3	0.1760	0.7180	0.0147	0.1640	-0.0611	-0.0715
Sous - modèle PLS - Mesures PQM Ligne C						
K = 3						
1	0.5350	0.5350	0.2720	0.2720	0.2540	0.2540
2	0.3510	0.8850	0.0345	0.3070	0.0286	0.2750

**Tableau B.1.1.2 Coefficients Centrés – Réduits et Valeurs VIP  
du modèle PLS multi-bloc.**

<b>Modèle</b>	<b>Composante</b>	<b>Coefficient CR</b>	<b>VIP [1]</b>
Copeaux	1	0.0854	1.5924
	2	0.0604	1.1263
	3	0.0321	0.5984
	4	0.0042	0.0790
	5	0.0084	0.1559
	6	0.0171	0.3195
Raffineur Primaire Ligne A	1	0.0966	1.8000
	2	0.0418	0.7794
	3	0.0191	0.3565
	4	0.0165	0.3069
	5	0.0115	0.2145
Raffineur Secondaire Ligne A	1	0.0944	1.7592
	2	0.0532	0.9908
	3	0.0222	0.4133
	4	0.0122	0.2282
PQM Ligne A1	1	0.0945	1.7606
	2	0.0178	0.3320
Raffineur Primaire Ligne B	1	0.0677	1.2614
	2	0.0354	0.6597
	3	0.0199	0.3707
	4	0.0188	0.3502
Raffineur Secondaire Ligne B	1	0.0601	1.1198
	2	0.0404	0.7535
	3	0.0221	0.4118
PQM Ligne B	1	0.0709	1.3221
	2	0.0357	0.6656
Raffineur Ligne C	1	0.0787	1.4663
	2	0.0470	0.8760
	3	0.0357	0.6649
PQM Ligne C	1	0.1175	2.1895
	2	0.0363	0.6775

Tableau B.1.1.3 Coefficients Centrés – Réduits et Valeurs VIP des sous - modèles.

Variable	Coefficient CR	VIP	Variable	Coefficient CR	VIP
<b>Copeaux</b>			<b>Raffineur Secondaire Ligne A</b>		
		<b>A = 6</b>			<b>A = 4</b>
Luminance	-0.0613	0.5986	Entrefer Plate	-0.1853	1.3268
Espèce 1	-0.1878	1.0720	Entrefer Conique	-0.0250	0.7096
Espèce 2	0.2146	1.6067	Âge des plaques	-0.1835	1.1832
Espèce 3	-0.0635	1.0773	Consistance	-0.0682	0.5970
Poids spécifique	0.0540	0.4139	Charge du moteur	0.1407	1.2429
Vrac	0.1702	1.5698	Énergie Spécifique	-0.0129	0.3673
Pin	0.1208	0.9806	Production Ligne A	0.1198	0.9216
1.125	0.0933	1.1577	Dilution alimentée / PR	0.0964	0.9857
0.875	-0.0359	1.0128	Dilution zone plate / PR	-0.0747	0.5653
0.625	-0.0402	0.5462	Dilution zone conique / PR	0.1101	0.8729
0.375	0.1489	1.2588	ES PA/ES SA	-0.2709	1.5484
0.1875	-0.1316	1.0953			
Fines	0.2134	1.1952			
Blancheur	0.1175	0.5891	<b>Mesures PQM Ligne A</b>		
Humidité	-0.0537	0.7997	Total de bûchettes	0.2127	1.1615
% Bûchettes	-0.1237	0.9560	Égouttage	-0.0862	0.2879
% Écorce	0.1201	0.7578	Longueur de fibres	0.2611	1.2523
% Pourris	0.0434	0.5320			
% Noeuds	0.0855	0.4092			
<b>Raffineur Primaire Ligne A1</b>			<b>Raffineur Primaire Ligne B</b>		
		<b>A = 5</b>			<b>A = 4</b>
Entrefer Plate	-7.25E-05	4.10E-01	Entrefer Plate	-0.1104	1.0172
Entrefer Conique	-3.60E-02	5.81E-01	Entrefer Conique	0.2729	1.8205
Âge des plaques	1.40E-01	5.77E-01	Âge des plaques	0.1664	0.6625
Consistance	1.34E-01	1.19E+00	Consistance	0.0563	0.9409
Charge du moteur	1.17E-01	5.14E-01	Charge du moteur	0.1680	0.7882
Énergie Spécifique	-2.43E-01	1.24E+00	Énergie Spécifique	0.2422	0.9825
Dilution alimentée / PR	-2.23E-01	1.65E+00	Dilution alimentée / PR	-0.2020	0.9699
Dilution zone plate / PR	1.16E-01	1.21E+00	Dilution zone conique / PR	-0.0417	0.7279
Dilution zone conique / PR	5.94E-02	6.59E-01	Dilution Soufflage / PR	-0.0716	0.5193
Dilution Soufflage / PR	6.08E-02	1.15E+00			

Tableau B.1.1.3 Coefficients Centrés – Réduits et Valeurs VIP des sous – modèles  
(continuation).

Variable	Coefficient CR	VIP	Variable	Coefficient CR	VIP
<b><i>Raffineur S2</i></b>			<b><i>Raffineur RJ</i></b>		
<i>A = 3</i>			<i>A = 3</i>		
Entrefer Plate	0.2087	1.2431	Entrefer Plate	0.0265	0.9284
Entrefer Conique	0.0775	0.8519	Entrefer Conique	0.0990	0.8738
Âge des plaques	-0.0355	0.3922	Âge des plaques	-0.1387	0.8222
Consistance	0.0190	0.6789	Consistance	0.0558	0.6417
Charge du moteur	0.1177	1.2299	Charge du moteur	-0.1374	0.8718
Énergie Spécifique	0.1400	1.2102	Énergie Spécifique	-0.0737	0.9831
Production Ligne A	-0.0018	0.6614	Production Ligne A	0.0004	0.7965
Dilution alimentée / PR	0.1851	1.2340	Dilution alimentée / PR	-0.0334	0.9044
Dilution zone conique / PR	-0.0340	1.0030	Dilution zone plate / PR	-0.1470	1.3598
ES PB/ES SB	0.0157	1.0822	Dilution zone conique / PR	-0.2510	1.5030
<b><i>Mesures PQM L2</i></b>			<b><i>Mesures PQM RJ</i></b>		
<i>A = 2</i>			<i>A = 2</i>		
Total de bûchettes	0.2031	1.0410	Total de bûchettes	0.2319	1.0082
Égouttage	-0.2000	0.8447	Égouttage	-0.2976	0.6929
Longueur de fibres	0.2254	1.0968	Longueur de fibres	0.3570	1.2262

## B.1.2 Modèles ayant comme variable de réponse l'indice de tension SM du papier.

Tableau B.1.2.1 Coefficients  $R^2X$ ,  $R^2Y$  et  $Q^2$  du modèle PLS multi-bloc et des sous modèles.

Modèle PLS Multi-bloc de l'Indice de Tension SM - 1ère période						
No. d'observations (N)=1651, No. de Variables Explicatives (K)=21						
A	$R^2X$	$R^2X(cum)$	$R^2Y$	$R^2Y(cum)$	$Q^2$	$Q^2(cum)$
1	0.0947	0.0947	0.384	0.384	0.378	0.378
2	0.0774	0.172	0.0555	0.439	0.0789	0.428
3	0.0623	0.234	0.0132	0.453	0.00772	0.432
Sous - modèle PLS – Propriétés des copeaux						
K = 4						
1	0.4140	0.4140	0.0508	0.0508	0.0496	0.0496
2	0.3420	0.7550	0.0089	0.0597	0.0085	0.0577
Sous - modèle PLS – Mesures PQM						
K = 6						
1	0.3490	0.3490	0.1880	0.1880	0.1870	0.1870
2	0.1360	0.4850	0.0133	0.2020	0.0119	0.1960
3	0.1540	0.6400	0.0040	0.2060	0.0029	0.1990
Sous - modèle PLS - Système d'approche						
K = 14						
1	0.2290	0.2290	0.0885	0.0885	0.0854	0.0854
2	0.1480	0.3760	0.0308	0.1190	0.0301	0.1130
3	0.0946	0.4710	0.0130	0.1320	0.0095	0.1210
4	0.0686	0.5400	0.0060	0.1380	-0.0016	0.1200
5	0.0576	0.5970	0.0076	0.1460	-0.0029	0.1170
6	0.0636	0.6610	0.0037	0.1500	-0.0027	0.1150
Sous - modèle - Machine à papier						
K = 27						
1	0.0802	0.0802	0.1850	0.1850	0.1750	0.1750
2	0.0692	0.1490	0.0491	0.2340	0.0445	0.2120
3	0.1080	0.2580	0.0110	0.2450	0.0017	0.2130
4	0.0829	0.3410	0.0085	0.2540	-0.0031	0.2110
5	0.0849	0.4260	0.0047	0.2580	-0.0079	0.2050
6	0.0485	0.4740	0.0018	0.2600	-0.0134	0.1940
7	0.0368	0.5110	0.0019	0.2620	-0.0144	0.1820
8	0.0369	0.5480	0.0027	0.2650	-0.0188	0.1670
9	0.0316	0.5790	0.0018	0.2670	-0.0172	0.1530
10	0.0247	0.6040	0.0047	0.2710	-0.0219	0.1340

**Tableau B.1.2.2 Coefficients Centrés – Réduits et Valeurs VIP  
du modèle PLS multi-bloc.**

<b>Modèle</b>	<b>Composante</b>	<b>Coefficient CR</b>	<b>VIP [3]</b>
Copeaux	1	0.1016	1.2185
	2	-0.0424	0.6811
PQM	1	0.3385	2.2420
	2	0.0329	0.6348
	3	0.0307	0.3974
Système d'approche	1	0.2400	1.5506
	2	0.0092	1.1151
	3	0.0824	0.5780
	4	0.0313	0.3031
	5	0.0855	0.4613
	6	0.0277	0.3201
Machine à papier	1	0.2678	2.1993
	2	0.2115	1.2281
	3	0.1779	0.9732
	4	-0.0247	0.8269
	5	0.0402	0.4541
	6	0.0200	0.3113
	7	0.0809	0.3242
	8	-0.0567	0.5159
	9	0.0949	0.3833
	10	0.0382	0.5271

Tableau B.1.2.3 Coefficients Centrés – Réduits et Valeurs VIP des sous - modèles.

Variable	Coefficient CR	VIP	Variable	Coefficient CR	VIP
<i>Copeaux</i>		<i>A = 2</i>	<i>Machine à papier</i>		<i>A = 10</i>
Luminance	-0.0216	0.1363	Pression CA	-0.0825	0.8132
Espèce 1 (%)	-0.1688	1.0559	Consistance CA	0.1357	1.1357
Espèce 2 (%)	0.1491	1.3727	Floculation CA	-0.0796	0.4437
Espèce 3 (%)	-0.0721	0.9911	Température CA	0.0807	0.9586
<i>Mesures PQM</i>		<i>A = 3</i>	Turbidité	0.1711	0.8622
LA Égouttage	0.0044	0.1352	Vide C. Transfert	0.0166	0.2808
LA Longueur de fibres	0.1249	1.3442	Vide 2 <sup>ème</sup> Caisse	0.0115	0.2031
LB Égouttage	0.0446	0.1432	Vide aspirant	0.1314	0.6977
LB Longueur de fibres	0.0459	1.1489	Vide 1ère presse	0.2811	1.7099
LC Égouttage	-0.1895	0.7866	Vide poussé 1ère presse	-0.0893	0.7927
LC Longueur de fibres	0.3124	1.4885	Const. eau blanche	-0.0042	0.3509
<i>Système d'approche</i>		<i>A = 6</i>	Vide réglage automatique	0.2376	1.3559
%Cassés	0.0771	0.7916	Ouverture de lèvres	-0.0199	0.9868
%DES	-0.2570	1.4109	Ratio vit. jet/toile	0.0374	0.4509
%PTM	0.1915	1.0543	Vitesse rouleau	-0.1253	0.5413
Égouttage mélange pâte	-0.0624	1.6006	Vitesse toile	0.0907	0.3540
Égouttage DES	0.0372	0.5195	Presse1 arrière	-0.0532	0.4250
Égouttage PTM	0.0666	0.8018	Press1 front	-0.0467	0.5696
pH mélange pâte	0.1183	0.9098	Press2 arrière	-0.0096	0.6042
pH PTM	0.0197	0.4772	Press2 front	0.0380	0.3079
pH DES	-0.0422	0.6321	Humidité	-0.0388	0.3238
Consist. mélange pâte	0.0293	0.1825	Degré Blancheur	-0.2519	2.3589
Blancheur PTM	0.2061	1.4708	Épaisseur	0.1699	1.0800
Blancheur DES	-0.0471	0.5830	% Tirage total	0.1972	0.6857
Ratio TMP/DIP	-0.1209	1.2408	Vapeur total / PR	0.1645	1.9614
Ag. rétention / PR	-0.0878	1.1249	Ag. blanch. / PR	0.0881	1.4889
			Ramasse pâte / PR	0.0750	1.0672



## B.1.3 Modèles ayant comme variable de réponse le degré de blancheur du papier.

Tableau B.1.3.1 Coefficients  $R^2X$ ,  $R^2Y$  et  $Q^2$  du modèle PLS multi-bloc et des sous modèles.

Modèle PLS Multi-bloc du Degré de Blancheur du papier - 1ère période						
Observations (N)=1611, No. de Variables Explicatives (K)=13						
A	$R^2X$	$R^2X(cum)$	$R^2Y$	$R^2Y(cum)$	$Q^2$	$Q^2(cum)$
1	0.1340	0.1340	0.2640	0.2640	0.2620	0.2620
2	0.1070	0.2400	0.0274	0.2910	0.0336	0.2870
Sous - modèle PLS - Propriétés de la pâte						
K = 12						
1	0.1460	0.1460	0.1830	0.1830	0.1780	0.1780
2	0.1610	0.3070	0.0361	0.2190	0.0406	0.2120
3	0.1510	0.4570	0.0033	0.2220	0.0020	0.2130
4	0.1000	0.5580	0.0009	0.2230	-0.0040	0.2100
5	0.0856	0.6430	0.0007	0.2240	-0.0048	0.2060
Sous - modèle PLS - Caisse d'arrivée						
K = 4						
1	0.3500	0.3500	0.1190	0.1190	0.1180	0.1180
2	0.1850	0.5350	0.0004	0.1190	-0.0002	0.1180
3	0.2660	0.8010	0.0000	0.1190	-0.0002	0.1180
Sous - modèle PLS - Adjuvants						
K = 2						
1	0.5740	0.5740	0.1340	0.1340	0.1320	0.1320
2	0.4260	1.0000	0.0016	0.1360	0.0001	0.1320
Sous - modèle PLS - Machine à papier						
K = 9						
1	0.1840	0.1840	0.0683	0.0683	0.0651	0.0651
2	0.1090	0.2930	0.0096	0.0779	0.0046	0.0694
3	0.1440	0.4370	0.0007	0.0786	-0.0027	0.0669
4	0.1000	0.5380	0.0003	0.0788	-0.0046	0.0626
5	0.0869	0.6250	0.0010	0.0798	-0.0049	0.0581

Tableau B.1.3.2 Coefficients Centrés – Réduits et Valeurs VIP  
du modèle PLS multi-bloc.

Modèle	Composante	Coefficient CR	VIP[2]
Propriétés de la Pâte	1	0.2910	2.2632
	2	0.2115	1.1736
	3	0.0449	0.3054
	4	0.0076	0.0455
	5	0.0722	0.6119
Caisse d'arrivée	1	0.1941	1.8555
	2	0.0067	0.1021
	3	0.0313	0.2268
Paper Machine	1	0.1775	1.3827
	2	0.1153	0.6458
	3	0.0183	0.1166
	4	-0.0390	0.4163
	5	0.0100	0.1034

Tableau B.1.3.3 Coefficients Centrés – Réduits et Valeurs VIP des sous - modèles.

Variable	Coefficient CR	VIP[5]	Variable	Coefficient CR	VIP[5]
<b>Propriétés de la pâte</b>			<b>Caisse d'arrivée</b>		
<i>A = 5</i>			<i>A = 3</i>		
%Cassés	-0.1898	1.3234	Pression CA	-0.2298	1.4816
%DES	0.1198	1.0704	Consistance CA	-0.0203	0.1251
%PTM	0.0295	0.4814	Floculation CA	-0.1436	1.1519
Égouttage mélange pâte	-0.0186	0.8874	Température CA	-0.0931	0.6799
Égouttage DES	0.1316	0.8140	<b>Paper Machine</b>		
Égouttage PTM	0.2703	1.8430			
pH mélange pâte	-0.1879	1.4018	Vide réglage automatique	0.0636	0.3990
pH PTM	-0.0389	0.3294	Vide C. Transfert	0.0048	0.0837
pH DES	-0.0652	0.5779	Vide 2 <sup>ème</sup> Caisse	0.0104	0.0716
Consist. mélange pâte	-0.0002	0.0521	Vide Aspirant	-0.0519	1.1196
Blancheur TMP	0.0455	0.5531	Vide 1ère presse	-0.1430	1.4344
Blancheur DES	0.2196	1.1439	Vide poussé 1ère presse	0.0079	1.2450
<b>Adjuvants</b>			<i>A = 2</i>		
Ag. rétention / PR	-0.0932	0.5547	% Tirage	0.0129	0.4634
Ag. blanch. / PR	-0.3409	1.3009	Vapeur total / PR	-0.1871	1.6887
			Ramasse pâte / PR	0.0857	0.9492

## B.1.4 Modèles ayant comme variable de réponse le débit total de vapeur.

Tableau B.1.4.1 Coefficients  $R^2X$ ,  $R^2Y$  et  $Q^2$  du modèle PLS multi-bloc et des sous modèles.

<b>Modèle PLS Multi-bloc du Débit Total de Vapeur - 1ère période</b>						
No. d'observations (N)=1223, No. de Variables Explicatives (K)=10						
A	$R^2X$	$R^2X(cum)$	$R^2Y$	$R^2Y(cum)$	$Q^2$	$Q^2(cum)$
1	0.1990	0.1990	0.1590	0.1590	0.1520	0.1520
2	0.1270	0.3260	0.0190	0.1780	0.0130	0.1630
Sous - modèle PLS - Propriétés de la pâte						
K = 12						
1	0.2450	0.2450	0.1240	0.1240	0.1180	0.1180
2	0.0945	0.3390	0.0194	0.1430	0.0091	0.1260
3	0.0792	0.4180	0.0094	0.1530	-0.0090	0.1190
4	0.0778	0.4960	0.0079	0.1600	-0.0137	0.1060
5	0.0309	0.5270	0.0072	0.1680	-0.0149	0.0932
6	0.1100	0.6370	0.0000	0.1680	-0.0093	0.0847
Sous - modèle PLS - Caisse d'arrivée						
K = 4						
1	0.3620	0.3620	0.0675	0.0675	0.0633	0.0633
2	0.2170	0.5800	0.0025	0.0700	-0.0028	0.0607
3	0.1920	0.7720	0.0006	0.0706	-0.0039	0.0571
Sous - modèle PLS - Adjuvants						
K = 2						
1	0.6	0.6	0.0447	0.0447	0.043	0.043

Tableau B.1.4.2 Coefficients Centrés – Réduits et Valeurs VIP  
du modèle PLS multi-bloc.

Modèle	Composante	Coefficient CR	VIP[2]
Pâte	1	0.254395	1.97109
	2	0.0840828	0.791003
	3	0.0741927	0.644111
	4	0.100305	0.660972
	5	0.084647	0.5473
	6	0.0879253	0.64035
Caisse d'arrivée	1	0.098176	1.51386
	2	0.0507255	0.321265
	3	0.0410452	0.2501
Adjuvants	1	0.0879329	1.21254

Tableau B.1.4.3 Coefficients Centrés – Réduits et Valeurs VIP des sous - modèles.

Variable	Coefficient CR	VIP
<b>Propriétés de la pâte</b>		<b>A = 6</b>
%Cassés	-0.0499	0.6497
%DES	-0.2675	1.4645
%PTM	0.3083	1.6222
Égouttage mélange pâte	-0.2769	1.7242
Égouttage DES	0.0414	0.5789
Égouttage PTM	-0.2468	1.2320
pH mélange pâte	0.0414	0.4092
pH PTM	-0.0012	0.0949
pH DES	-0.0704	0.7083
Consist. mélange pâte	0.0018	0.1469
Blancheur TMP	0.1613	0.9808
Blancheur DES	-0.0507	0.5605
<b>Caisse d'arrivée</b>		<b>A = 3</b>
Pression CA	0.1933	1.4578
Consistance CA	-0.0505	0.4320
Floculation CA	0.1145	1.1890
Température CA	-0.0055	0.5241
<b>Adjuvants</b>		<b>A = 1</b>
Ag. rétention / PR	0.1353	1.0268
Ag. blanch. / PR	0.1281	0.9724

## B.2 Deuxième Période (Octobre 2003 à Janvier 2004).

### B.2.1 Modèles ayant comme variable de réponse l'indice de tension des formettes.

Tableau B.2.1.1 Coefficients  $R^2X$ ,  $R^2Y$  et  $Q^2$  du modèle PLS multi-bloc et des sous modèles.

<b>Modèle PLS Multi-bloc de l' Indice de Tension des Formettes - 2ème période</b>						
No. d'observations (N)=115, No. de variables (K)=33						
A	$R^2X$	$R^2X(cum)$	$R^2Y$	$R^2Y(cum)$	$Q^2$	$Q^2(cum)$
1	0.1080	0.1080	0.4270	0.4270	0.2810	0.2810
Sous - modèle PLS - Propriétés des Copeaux						
K = 20						
1	0.1180	0.1180	0.1140	0.1140	-0.1050	-0.1000
2	0.1170	0.2350	0.0763	0.1910	-0.0964	-0.2060
3	0.1650	0.4000	0.0170	0.2080	-0.1020	-0.3270
4	0.0970	0.4970	0.0281	0.2360	-0.1750	-0.4590
5	0.0769	0.5740	0.0209	0.2570	-0.1330	-0.6050
6	0.0652	0.6390	0.0106	0.2670	-0.1690	-0.7660
Sous - modèle PLS - Raffineur P1						
K = 10						
1	0.1450	0.1450	0.0631	0.0631	-0.1190	-0.1000
2	0.2070	0.3520	0.0119	0.0750	-0.0984	-0.2080
3	0.2120	0.5640	0.0012	0.0762	-0.0956	-0.3240
4	0.0754	0.6390	0.0025	0.0787	-0.1270	-0.4560
Sous - modèle PLS - Raffineur S1						
K = 11						
1	0.2350	0.2350	0.1040	0.1040	-0.0134	-0.0134
2	0.0972	0.3320	0.0479	0.1520	-0.2000	-0.1150
3	0.1430	0.4750	0.0153	0.1670	-0.1450	-0.2260
4	0.1700	0.6450	0.0050	0.1720	-0.1310	-0.3490
Sous - modèle PLS - Mesures PQM L1						
K = 3						
1	0.3410	0.3410	0.0781	0.0781	0.0137	0.0137
2	0.3620	0.7030	0.0066	0.0847	-0.0359	-0.0216

Tableau B.2.1.1 Coefficients  $R^2X$ ,  $R^2Y$  et  $Q^2$  du modèle PLS multi-bloc  
et des sous modèles (continuation).

A	$R^2X$	$R^2X(cum)$	$R^2Y$	$R^2Y(cum)$	$Q^2$	$Q^2(cum)$
Sous - modèle PLS - Raffineur P2						
K = 10						
1	0.3210	0.3210	0.1530	0.1530	0.1160	0.1160
2	0.1280	0.4490	0.0828	0.2350	-0.0016	0.1140
3	0.1100	0.5590	0.0231	0.2590	-0.0601	0.0609
4	0.0909	0.6500	0.0043	0.2630	-0.0815	-0.0156
Sous - modèle PLS - Raffineur S2						
K = 11						
1	0.1420	0.1420	0.0552	0.0552	-0.1660	-0.1000
2	0.1590	0.3020	0.0055	0.0606	-0.1530	-0.2100
3	0.1440	0.4460	0.0046	0.0652	-0.1450	-0.3310
4	0.1180	0.5640	0.0054	0.0707	-0.1330	-0.4640
5	0.1450	0.7090	0.0024	0.0731	-0.1250	-0.6110
Sous - modèle PLS - Mesures PQM L2						
K = 3						
1	0.4710	0.4710	0.1670	0.1670	0.1070	0.1070
2	0.3090	0.7800	0.0327	0.2000	0.0003	0.1080
Sous - modèle PLS - Raffineur RJ						
K = 11						
1	0.293	0.293	0.0199	0.0199	-0.0402	-0.0402
2	0.114	0.407	0.00883	0.0288	-0.0841	-0.128
3	0.184	0.591	0.0125	0.0413	-0.098	-0.238
4	0.166	0.758	0.00627	0.0475	-0.12	-0.362
Sous - modèle PLS - Mesures PQM RJ						
K = 3						
1	0.498	0.498	0.101	0.101	0.0435	0.0435
2	0.305	0.803	0.0122	0.113	-0.0429	0.00244

Tableau B.2.1.2 Coefficients Centrés – Réduits et Valeurs VIP  
du modèle PLS multi-bloc.

Modèle	Composante	Coefficient CR	VIP[1]
Copeaux	1	0.1094	1.7023
	2	0.0867	1.3491
	3	0.0372	0.5790
	4	0.0578	0.8996
	5	0.0528	0.8213
	6	0.0398	0.6190
Raffineur Primaire Ligne A	1	0.0816	1.2697
	2	0.0354	0.5517
	3	0.0117	0.1825
	4	0.0192	0.2996
Raffineur Secondaire Ligne A	1	0.1069	1.6648
	2	0.0723	1.1260
	3	0.0430	0.6696
	4	0.0233	0.3633
PQM Ligne A	1	0.0899	1.3995
	2	0.0244	0.3797
Raffineur Primaire Ligne B	1	0.1234	1.9215
	2	0.0911	1.4178
	3	0.0510	0.7939
	4	0.0218	0.3392
Raffineur Secondaire Ligne B	1	0.0742	1.1551
	2	0.0220	0.3432
	3	0.0198	0.3079
	4	0.0221	0.3447
	5	0.0152	0.2373
PQM Ligne B	1	0.1304	2.0302
	2	0.0565	0.8793
Raffineur Ligne C	1	0.0444	0.6904
	2	0.0408	0.6347
	3	0.0433	0.6747
	4	0.0361	0.5625
PQM Ligne C	1	0.0995	1.5486
	2	0.0290	0.4512

Tableau B.2.1.3 Coefficients Centrés – Réduits et Valeurs VIP des sous – modèles.

Variable	Coefficient CR	VIP	Variable	Coefficient CR	VIP
<b>Copeaux</b>			<b>Raffineur S1</b>		
		<b>A = 6</b>			<b>A = 4</b>
Luminance	-0.0230	0.6275	Entrefer Plate	-0.0188	0.3826
Espèce 1	-0.0675	0.8801	Entrefer Conique	0.0694	0.7113
Espèce 2	0.0276	0.6450	Âge des plaques	-0.1060	1.4395
Espèce 3	0.0217	0.8724	Consistance	0.0509	0.9325
Poids spécifique	-0.3263	1.4544	Charge du moteur	0.2798	1.0012
Vrac	-0.0649	0.6418	Énergie Spécifique	-0.2910	1.1703
Pin	-0.0766	0.7881	Production Ligne A	-0.0680	0.5150
1.125	0.1061	0.8723	Dilution alimentée / PR	0.1428	1.0273
0.875	0.1578	1.1963	Dilution zone plate / PR	-0.0420	1.0753
0.625	-0.0332	1.0995	Dilution zone conique / PR	-0.0598	1.1575
0.375	-0.0784	0.7288	Dilution Soufflage / PR	0.2050	1.1033
0.1875	0.5291	1.3272			
Fines	-0.1981	1.2031			
Blancheur	-0.0441	0.9328	<b>Mesures PQM L1</b>		
Humidité	-0.3888	1.6071	Total de bûchettes	0.0382	0.5250
% Bûchettes	0.1427	0.9740	Égouttage	-0.1286	0.5635
% Écorce	0.2015	0.8122	Longueur de fibres	0.2665	1.5514
% Pourris	-0.0110	0.7783			
% Noeuds	0.1008	1.0984			
Petits	-0.0225	0.7013			
<b>Raffineur P1</b>			<b>Raffineur P2</b>		
		<b>A = 4</b>			<b>A = 4</b>
Entrefer Plate	-0.1174	0.7284	Entrefer Plate	-0.2729	1.2519
Entrefer Conique	0.2005	1.7895	Entrefer Conique	0.1790	0.6985
Âge des plaques	-0.0842	1.0091	Âge des plaques	-0.2779	1.2058
Consistance	0.0005	0.6032	Consistance	0.1168	0.9274
Charge du moteur	-0.0090	0.4273	Charge du moteur	-0.0828	1.0248
Énergie Spécifique	-0.0090	0.5789	Énergie Spécifique	0.1033	1.1259
Dilution alimentée / PR	-0.0739	0.7284	Dilution alimentée / PR	-0.0245	0.9700
Dilution zone plate / PR	0.0742	0.8409	Dilution zone plate / PR	0.0561	0.3903
Dilution zone conique / PR	0.0697	0.3524	Dilution zone conique / PR	0.2134	1.0929
Dilution Soufflage / PR	0.1899	1.7337	Dilution Soufflage / PR	-0.2892	1.0127



Tableau B.2.1.3 Coefficients Centrés – Réduits et Valeurs VIP des sous – modèles  
(continuation).

Variable	Coefficient CR	VIP	Variable	Coefficient CR	VIP
<b><i>Raffineur S2</i></b>			<b><i>Raffineur RJ</i></b>		
		<b><i>A = 5</i></b>			<b><i>A = 4</i></b>
Entrefer Plate	0.0971	0.5457	Entrefer Plate	0.1253	1.0165
Entrefer Conique	-0.1890	1.4141	Entrefer Conique	-0.0339	0.5747
Âge des plaques	0.1190	1.1991	Âge des plaques	0.1486	1.5654
Consistance	0.0797	0.7552	Consistance	0.1162	1.0730
Charge du moteur	0.1453	0.7508	Charge du moteur	-0.0155	0.6960
Énergie Spécifique	-0.1038	0.4836	Énergie Spécifique	0.0683	0.8286
Production Ligne B	-0.2013	1.7569	Production Ligne C	0.1137	1.1791
Dilution alimentée / PR	0.0739	1.0474	Dilution alimentée / PR	0.1037	1.0767
Dilution zone plate / PR	0.0172	0.6276	Dilution zone plate / PR	0.0516	0.6550
Dilution zone conique / PR	-0.0173	0.6197	Dilution zone conique / PR	-0.0483	1.0813
Dilution Soufflage / PR	-0.0301	0.9670	Dilution Soufflage / PR	0.0245	0.8462
<b><i>Mesures PQM L2</i></b>			<b><i>Mesures PQM RJ</i></b>		
		<b><i>A = 2</i></b>			<b><i>A = 2</i></b>
Total de bûchettes	0.0778	0.5864	Total de bûchettes	0.1334	0.8685
Égouttage	-0.0940	0.6904	Égouttage	-0.1284	0.5483
Longueur de fibres	0.4459	1.4763	Longueur de fibres	0.2872	1.3946

## B.2.2 Modèles ayant comme variable de réponse l'indice de tension SM du papier.

Tableau B.2.2.1 Coefficients  $R^2X$ ,  $R^2Y$  et  $Q^2$  du modèle PLS multi-bloc et des sous modèles.

<b>Modèle PLS Multi-bloc de l'Indice de Tension SM - 2ème période</b>						
No. d'observations (N)=1935, No. de variables explicatives (K)=20						
A	$R^2X$	$R^2X(cum)$	$R^2Y$	$R^2Y(cum)$	$Q^2$	$Q^2(cum)$
1	0.0933	0.0933	0.4760	0.4760	0.4700	0.4700
2	0.0586	0.1520	0.0217	0.4980	0.0218	0.4810
Sous - modèle PLS - Propriétés des copeaux						
K = 3						
1	0.4720	0.4720	0.0092	0.0092	0.0086	0.0086
2	0.5280	1.0000	0.0013	0.0105	0.0012	0.0098
Sous - modèle PLS - Mesures PQM						
K = 6						
1	0.3280	0.3280	0.1080	0.1080	0.1060	0.1060
2	0.1340	0.4620	0.0289	0.1370	0.0286	0.1320
3	0.1710	0.6340	0.0034	0.1410	0.0020	0.1330
Sous - modèle PLS - Système d'approche						
K = 14						
1	0.2010	0.2010	0.1730	0.1730	0.1700	0.1700
2	0.1140	0.3150	0.0519	0.2250	0.0572	0.2180
3	0.1000	0.4160	0.0098	0.2350	0.0049	0.2210
4	0.0940	0.5100	0.0033	0.2380	-0.0033	0.2190
5	0.0667	0.5760	0.0018	0.2400	-0.0053	0.2150
6	0.0693	0.6460	0.0008	0.2410	-0.0066	0.2100
Sous - modèle PLS - Machine à papier						
K = 27						
1	0.1570	0.1570	0.3500	0.3500	0.3450	0.3450
2	0.0714	0.2280	0.0796	0.4290	0.1100	0.4180
3	0.0660	0.2940	0.0209	0.4500	0.0205	0.4300
4	0.0516	0.3450	0.0121	0.4620	0.0001	0.4300
5	0.0860	0.4310	0.0041	0.4660	-0.0068	0.4260
6	0.0772	0.5090	0.0026	0.4690	-0.0042	0.4230
7	0.0510	0.5600	0.0015	0.4700	-0.0122	0.4160
8	0.0382	0.5980	0.0017	0.4720	-0.0165	0.4070
9	0.0284	0.6260	0.0012	0.4730	-0.0154	0.3980

**Tableau B.2.2.2 Coefficients Centrés – Réduits et Valeurs VIP  
du modèle PLS multi-bloc.**

<b>Modèle</b>	<b>Composante</b>	<b>Coefficient CR</b>	<b>VIP [2]</b>
Copeaux	1	0.0273	0.4934
	2	0.0412	0.2129
PQM	1	0.1497	1.5614
	2	0.0745	0.8076
	3	-0.0053	0.3908
Système d'approche	1	0.2148	1.9654
	2	0.0854	1.0719
	3	0.0273	0.4799
	4	0.0146	0.2667
	5	0.0235	0.1625
	6	-0.0219	0.2665
Machine à papier	1	0.3642	2.7856
	2	0.2252	1.3922
	3	0.1236	0.7077
	4	0.1410	0.7191
	5	0.0562	0.4226
	6	0.0216	0.2270
	7	0.0307	0.2284
	8	0.0095	0.1891
	9	0.0273	0.1539

Tableau B.2.2.3 Coefficients Centrés – Réduits et Valeurs VIP des sous - modèles.

Variable	Coefficient CR	VIP	Variable	Coefficient CR	VIP
<i>Copeaux</i>		<i>A = 2</i>	<i>Machine à papier</i>		<i>A = 9</i>
Espèce 1 (%)	-0.0937	1.5080	Pression CA	0.0261	0.6626
Espèce 2 (%)	0.0424	0.5765	Consistance CA	-0.1523	1.7911
Espèce 3 (%)	0.0136	0.6274	Floculation CA	-0.0405	0.5580
<i>Mesures PQM</i>		<i>A = 3</i>	Température CA	0.0189	0.7998
			Vide réglage automatique	0.0229	1.7484
LA Égouttage	0.0185	0.3501	Vide C. Transfert	0.1443	0.9157
LA Longueur de fibres	-0.0717	0.9515	Vide 2 <sup>ème</sup> Caisse	0.0015	0.1038
LB Égouttage	-0.0941	0.3559	Vide Aspirant	-0.0448	0.6679
LB Longueur de fibres	0.3783	1.8997	Vide 1ère presse	0.1352	1.0554
LC Égouttage	0.0183	0.2468	Vide poussé 1ère presse	0.1004	0.3250
LC Longueur de fibres	0.0239	1.0843	Const. eau blanche	-0.0242	1.0961
<i>Système d'approche</i>		<i>A = 6</i>	Turbidité	0.0529	0.3719
			Ouverture lèvres	-0.0469	1.5937
%Cassés	-0.0398	0.6130	Ratio vit. jet/toile	-0.0449	0.7584
%DES	-0.1201	1.6856	Vitesse rouleau	-0.0497	0.7342
%PTM	0.1418	1.4946	Vitesse toile	0.0020	0.6680
Égouttage mélange pâte	-0.1262	0.6267	Press1 arrière	0.1061	0.7951
Égouttage DES	0.1592	0.7603	Press1 front	0.1698	1.1771
Égouttage PTM	-0.2171	1.3856	Press2 arrière	0.4089	1.7196
pH mélange pâte	-0.0528	0.5279	Press2 front	0.0323	0.6532
pH PTM	0.0273	0.1994	Humidité	-0.0475	0.3766
pH DES	-0.0487	0.5857	Degré de blancheur	-0.0019	0.2571
Consist. mélange pâte	0.0044	0.2027	Épaisseur	0.2595	1.6549
Blancheur PTM	0.0825	0.8674	% Tirage	-0.2421	1.3397
Blancheur DES	-0.0190	0.2306	Ag. blancheur / PR	0.1451	0.9396
Ratio TMP/DIP	0.1937	1.6501	Vapeur total / PR	0.0289	0.5847
Ag. rétention / PR	-0.2231	1.1947	Ramasse pâte / PR	-0.0312	0.1509

### B.2.3 Modèles ayant comme variable de réponse la consistance de la caisse d'arrivée.

Tableau B.2.3.1 Coefficients  $R^2X$ ,  $R^2Y$  et  $Q^2$  du modèle PLS multi-bloc et des sous modèles.

<b>Modèle PLS Multiblock de la Consistencia de la caisse d'arrivée - 2ème période</b>						
No. d'bservations (N)=1975, No. de Variables Explicatives (K)=9						
A	$R^2X$	$R^2X(cum)$	$R^2Y$	$R^2Y(cum)$	$Q^2$	$Q^2(cum)$
1	0.1110	0.1110	0.3750	0.3750	0.3710	0.3710
Sous - modèle PLS - Propriétés de la pâte						
K = 12						
1	0.0949	0.0949	0.3320	0.3320	0.3270	0.3270
2	0.1370	0.2320	0.0320	0.3640	0.0442	0.3560
3	0.0923	0.3240	0.0031	0.3670	-0.0014	0.3550
4	0.0571	0.3810	0.0017	0.3680	-0.0041	0.3530
5	0.1480	0.5290	0.0015	0.3700	-0.0027	0.3510
6	0.0694	0.5980	0.0022	0.3720	-0.0019	0.3500
7	0.0801	0.6780	0.0018	0.3740	-0.0036	0.3470
Sous - modèle PLS – Adjuvants						
K = 2						
1	0.5160	0.5160	0.0032	0.0032	0.0027	0.0027
2	0.4840	1.0000	2.77E-05	0.0033	-0.0002	0.0025

Tableau B.2.3.2 Coefficients Centrés – Réduits et Valeurs VIP du modèle PLS multi-bloc.

Modèle	Composante	Coefficient CR	VIP[1]
Propriétés de la pâte	1	0.5829	2.8371
	2	0.1706	0.8303
	3	0.0506	0.2461
	4	0.0448	0.2180
	5	0.0305	0.1483
	6	0.0397	0.1930
	7	0.0248	0.1205
Adjuvants	1	0.0577	0.2806
	2	0.0053	0.0260

Tableau B.2.3.3 Coefficients Centrés – Réduits et Valeurs VIP des sous - modèles.

Variable	Coefficient CR	VIP
<b>Propriétés de la pâte</b>		<b>A = 7</b>
%Cassés	0.0457	0.2551
%DES	0.0584	0.5501
%PTM	-0.0905	0.6493
Égouttage mélange pâte	0.1714	0.6726
Égouttage DES	-0.3579	1.2495
Égouttage PTM	0.4954	2.7950
pH mélange pâte	-0.0287	0.1095
pH PTM	-0.0124	0.0719
pH DES	0.1995	1.0077
Consist. mélange pâte	0.0203	0.2632
Blancheur TMP	-0.0839	0.3761
Blancheur DES	0.0075	0.3767
<b>Adjuvants</b>		<b>A = 2</b>
Ag. blancheur / PR	-0.0281	0.8304
Ag. rétention / PR	0.0481	1.1448